CONTENTS

LIST OF CONTRIBUTORS vii

INTRODUCTION ix
  David Nevo

PART 1: CONCEPTUAL PERSPECTIVES

DIALOGUE EVALUATION: COMBINING INTERNAL AND EXTERNAL EVALUATION
  David Nevo

SCHOOL SELF-EVALUATION IN A DEMOCRACY
  Helen Simons

SCHOOL SELF-EVALUATION: ORIGINS, DEFINITION, APPROACHES, METHODS AND IMPLEMENTATION
  Jaap Scheerens

PART 2: CASE STUDIES

SCHOOL-BASED EVALUATION IN NORWAY: WHY IS IT SO DIFFICULT TO CONVINCE TEACHERS OF ITS USEFULNESS?
  Lars Monsen

EVALUATION AS SCHOOL IMPROVEMENT: A DEVELOPMENTAL PERSPECTIVE FROM ENGLAND
  David Hopkins and Mel West

SCHOOL SELF-EVALUATION IN THE NETHERLANDS
  Jaap Scheerens and Maria Hendriks

v
SELF-EVALUATION AS ENDEAVOUR TO UNDERSTAND QUALITY: EXPERIENCES IN AUSTRIAN SCHOOL-BASED EVALUATION
   Michael Schratz 145

INTERNAL AND EXTERNAL SCHOOL EVALUATION IN THE SPANISH EDUCATIONAL REFORM
   Alejandro Tiana 167

EVALUATING ACCELERATED SCHOOLS IN THE UNITED STATES
   Pia Lindquist Wong 187

SCHOOL-BASED EVALUATION AND THE NEO-LIBERAL AGENDA IN CANADA
   Les McLean 201

ISRAEL: INTERNAL SELF-EVALUATION AS A MEANS FOR GENERATING CHANGE
   Eiel Friedman and Pnina Golan-Cook 219

SCOTLAND: SCHOOLS SPEAKING FOR THEMSELVES
   John MacBeath 243

SCHOOL EVALUATION IN GERMANY: A MEANS TO IMPROVE SCHOOL QUALITY
   Claus G. Buhren 261
LIST OF CONTRIBUTORS

Claus G. Buhren
University of Dortmund,
Germany

Etel Friedman
Tel Aviv University,
Israel

Pnina Golan-Cook
The Hebrew University,
Israel

Maria Hendriks
University of Twente,
The Netherlands

David Hopkins
The University of Nottingham,
England

Pia Lindquist Wong
California State University,
Sacramento, USA

John MacBeath
University of Cambridge,
England

Les McLean
University of Toronto,
Canada

Lars Monsen
Lillehammer College,
Norway

David Nevo
School of Education,
Tel Aviv University, Israel

Jaap Scheerens
University of Twente,
The Netherlands

Michael Schratz
University of Innsbruck,
Austria
LIST OF CONTRIBUTORS

Helen Simons
University of Southampton, England

Alejandro Tiana
The Open University, Spain

Mel West
The University of Nottingham, England
INTRODUCTION

Evaluation has always existed within the school, mainly in its classrooms, although in a very limited capacity. It was there even before administrators and policy makers started to use it for decision making and accountability, and curriculum developers were using it to improve newly developed instructional materials. But evaluation within the school was usually limited to student evaluation or student assessment, conducted by means of tests and examinations, and reported in the form of grades and report cards. Systematic evaluation of instruction, program evaluation, evaluation of teachers and school administrators, as well as the evaluation of the school as a whole – were rarely a significant component in the life of schools.

At the same time much has changed in educational evaluation within the educational system. Since the early sixties, much progress has been made in defining the meaning of educational evaluation, identifying its functions, developing multiple methodologies for its conduct, and understanding the variables that might affect its utilization. But most of these happened in the context of curriculum development and the evaluation of large-scale projects. Many new evaluation concepts were developed – mainly in the United States – but only very few of them ever reached the school building. The most famous concept among those that did reach the classroom and the school building was probably the distinction between Formative Evaluation and Summative Evaluation. This distinction, originally suggested in the context of curriculum evaluation, is now well known to teachers around the world, although sometimes misinterpreted.

Many other concepts and practices, common in program evaluation, remained far away from schools and teachers. Among them, evaluation criteria, evaluation objects, variables of evaluation, evaluation audiences and stakeholders, or the distinction between description and judgement – just to mention a few. Most schools are not yet used to evaluate their programs and projects. Teachers do not develop their tests in a systematic way, nor do they assess the quality of their tests or other measures of achievement. School principals seldom exploit the current knowledge on teacher evaluation methods to evaluate their teachers, and hardly ever do they develop data pools, comprised of carefully collected data, that could be used for decision making and accountability. Not too many schools know how to face external evaluation with findings of their own internal
evaluation, rather than trying to undermine its credibility or question its legitimacy.

But there are some exceptions in schools that do use evaluation in a broader perspective than student assessment. Schools that have their internal evaluation teams; schools that evaluate systematically their programs to improve their quality; schools that develop school portfolios to demonstrate their quality and identify areas that need to be improved. These schools are scattered around the world in many educational systems. They present a very important advance in program evaluation, attesting to its relevancy to schools and teachers, and demonstrating how program evaluation can be applied at the school level. A series entitled Advances in Program Evaluation is an appropriate framework to learn from their experience.

This volume is comprised of thirteen chapters, divided into two parts. In the First Part, we discuss the concept of school-based evaluation from three perspectives. Nevo’s chapter presents a perception of school-based evaluation as a dialogue between internal and external evaluation. Simons’ chapter discusses school evaluation from a perspective of institutional self-evaluation in a democracy. Jaap Scheerens’s concludes this part discussing issues of definition, methods and implementation.

In the Second Part we present ten case studies from ten countries around the world: Norway, England, The Netherlands, Austria, Spain, United States, Canada, Israel, Scotland and Germany. All case studies are based on actual experience with school-based evaluation in various educational and social contexts with a wide range of local constraints, and reflecting multiple evaluation perspectives. Authors describe their educational context, identify their evaluation perspective, describe their specific school-based experience, point out difficulties they have encountered in their work, discuss its implications and make recommendations for further development of the concept of school-based evaluation and its practice.

This volume presents theory and practice from around the world suggesting that school-based evaluation is a “real thing” being practised in many educational systems. Not without problems but with many prospects. We do not suggest internal/self school evaluation as an alternative to external “objective” evaluation. We do try, however, to advocate the combination of both for the benefit of school accountability and school improvement and present some (convincing, we think) examples, from a variety of educational systems, showing that it can be done and it’s worth doing. We hope that you will find at least one of them relevant to your own educational system.

David Nevo
Editor
DIALOGUE EVALUATION: COMBINING INTERNAL AND EXTERNAL EVALUATION*

David Nevo

INTRODUCTION

Internal evaluation and external evaluation have long been under discussion in the evaluation literature (Scriven, 1967, 1991; Stufflebeam et al., 1971; Love, 1991; Mathison, 1991; Sonnichsen, 2000). Scriven (1991) gave the following definitions for internal and external evaluation in the context of program evaluation:

Internal evaluators (or evaluations) are those done by project staff, even if they are special evaluation staff – that is, even if they are external to the production/writing/service part of the project (p. 197).

An external evaluator is someone who is at least not on the project or program’s staff, or someone – in the case of personnel evaluation – other than the individual being evaluated, or their family or staff... It is best to regard externality as a continuum along which one tries to score as high as possible... (p. 159).

At the school level, internal evaluation can be performed by a teacher or a group of teachers, by other members of the school’s professional personnel, by the principal or other school administrators, or by a special staff member designated by the school to serve as a ‘school evaluator’. An external evaluation

* This chapter has been previously published in Studies in Educational Evaluation, Vol. 27(2), 2001.

School-Based Evaluation: An International Perspective, Volume 8, pages 3-16.
Copyright © 2002 by Elsevier Science Ltd.
All rights of reproduction in any form reserved.
ISBN: 0-7623-0861-3
of the school can be performed by the school district, the state department of education, or a ministry of education, using professional evaluators, regional inspectors, or a district/state/national evaluation department. An external evaluation of the school could also be conducted by an independent evaluation consultant or evaluation firm, commissioned by the school itself or its governing board.

EXTERNAL SCHOOL EVALUATION

For many years and in many countries school evaluation was tantamount to external evaluation. In some countries it was done mainly by inspectors (e.g. in the U.K. or other European countries), in some (e.g. in the USA) it was done by means of state or district assessment programs. Many educational systems combined both student assessment programs and overall school reviews, some of them conducted in a very systematic way by central units in the educational system (e.g. OFSTED in the U.K. or OER in New Zealand). The idea of accountability, which has been around for over a quarter century, and relatively newer ideas of setting standards and benchmarks as major means for school improvement (Wilson, 1996), can all be traced back to this long tradition of controlling schools by means of external evaluation.

Even before the term accountability was used, there was a clear demand by politicians, administrators, parents and the public at large that schools be evaluated externally to find out if they were fulfilling their duties. There was also a hope that such external evaluations would motivate teachers and school principals to work harder to improve their schools. This was true for democratic and non-democratic societies, and for centralized and decentralized educational systems alike. The demand never ceased, even when external evaluation was highly criticized by innovative educators and when internal evaluation was encouraged by way of an alternative.

INTERNAL EVALUATION

Parallel to the almost universal phenomenon of external school evaluation, many countries have more recently tended to apply newly developed evaluation methods at the school level in the form of internal evaluation or self-evaluation. Participatory Evaluation (Cousins & Whitmore, 1998), Empowerment Evaluation (Fetterman et al., 1996) Total Quality Management (TQM) and Action Research, are some attempts made to apply internal evaluation methods at the school level. These seem to be in line with other prevailing ideas such as reflection (Schon, 1983) and professionalization of teachers (Darling-Hamond, 1992) and
educational administrators/leaders. Reflection is, in a way, one kind of self-
evaluation although its proponents do not sufficiently stress the important role
of systematic data collection as a basis for reflection. A professional perception of
teaching regards teaching as a complex undertaking, suggesting that teachers
should identify needs, analyze goals, choose instructional strategies, and plan
and monitor their work. Evaluation becomes an integral part of the teaching
profession, relevant to various aspects of teachers’ responsibilities and no longer
narrowly limited to evaluating student achievements.

Internal evaluation is also a very important component of schools in
decentralized or decentralizing educational systems, which are following models
of school-based management or autonomous schools (Nevo, 1997). Autonomous
schools are urged to define their own educational aims, to be in charge of the
educational process and to evaluate their actions. With the expansion of school
authority, schools are also expected to take greater responsibility and to be
accountable for their deeds. These demands have emphasized the importance
of internal self-evaluation not only to support improvement but also to respond
to the school accountability requirement.

Internal evaluation becomes a major tool for school-based management,
serving decision making at various administrative levels. The improvement of
decision-making processes is especially important within the broader context
of ongoing decentralization and school empowerment, mentioned above.
Schools are granted significant authority and are, in return, expected to make
decisions autonomously. Internal self-evaluation is highly salient in such
situations: it can provide schools with means to improve decision-making
processes and make them more effective. An internal evaluator is usually better
acquainted with the local context of the evaluation and less threatening to those
being evaluated. S/he knows the local problems, communicates better with
the local people, and remains on site to facilitate the implementation of the
evaluation recommendations. Developing an internal evaluation mechanism
in a school is also an investment in an enduring resource for serving the
information needs of the school by means of data pools and school portfolios.

Internal evaluation is also an expression of school empowerment and transfer
of authority from the center to the periphery, from the central government to
the local community. Participation in the evaluation process may contribute
to the empowerment of an organization as a whole, as it acquires the ability to
monitor itself in a more systematic way and by gaining greater confidence in
its educational direction. Participation in the evaluation process can furthermore
empower individuals in the organization by providing them with evaluation
skills, which they can later apply in various contexts. The principals and the
teachers participating in evaluation activities can apply the newly acquired
evaluation skills and knowledge in other areas of their work. For example, teachers can use their knowledge of research methodology to teach pupils to perform investigative tasks as part of their schoolwork.

Another benefit of participation in evaluation activities is that it may increase teachers' involvement in decision-making processes outside the classroom, it may foster collegiality and collaboration amongst the teachers and serve as a means of promoting reflection. All these are central to development of teacher professionalization (Darling-Hammond, 1992).

However, pointing out the importance of internal evaluation or self-evaluation and its relevancy to professional teachers and autonomous schools should not lead us to ignore external evaluation. Control and accountability requirements and the right of the public to know in democratic societies have also to be kept in mind. Although internal evaluation, by providing information to parents and the community at large, can also enhance accountability, the credibility of its findings might be limited without external evaluation.

MUTUAL BENEFITS

If internal and external evaluation are both important, then the question is: Can they live together? I would like to suggest a positive answer to this question. They can coexist because we need both and because they might even benefit from each other.

How Can Internal Evaluation Benefit from External Evaluation?

At least in three ways. External evaluation can stimulate internal evaluation, it can expand its scope and legitimize its validity. Let's look in more detail into those three things.

1. Stimulating Internal Evaluation
Evaluation is a demanding undertaking whose benefits have still to be proved. It requires a significant amount of resources, such as funds, time and skilled personnel. We evaluators believe in its potential usefulness, but our research on evaluation utilization has so far yielded only meager support to our beliefs, and mainly in relation to conceptual use of evaluation (Cousins & Leithwood, 1986; Shulha & Cousins, 1997). The common wisdom is that evaluation is a required activity and hopefully a useful one.

It would be difficult to argue that doing self-evaluation comes naturally to teachers and schools, even less so – being evaluated by others. Schools will
surrender with or without protest to inevitable external evaluation requirements, imposed on them by the educational system. In some cases they might opt for internal self-evaluation if they believe that this will release them from external evaluation. Experience shows that schools might be motivated to engage in self-evaluation if faced with an external evaluation requirement, even when internal evaluation is not suggested as an alternative to external evaluation but only as a prior condition and counterpart. Some schools will tend to do internal evaluation to help them confront the external evaluation. Others believe, as I said already, that they might be released from the burden of external evaluation if they do internal evaluation or just establish an internal evaluation team.

Sometimes it's difficult to avoid the notion that the most important function of external evaluation is to motivate people and organizations to do internal evaluation . . .

2. Expanding the Scope of Internal Evaluation
Although internal evaluation usually has the advantage of being more sensitive to the local context of schools and their unique characteristics, it might suffer from a narrow perspective on their overall qualities. External evaluation can add commonalties to the uniqueness of the school and also provide a basis to judge its qualities. School inspectors can provide useful observations from their inspection region. Other external evaluators, such as professional state evaluation teams, may be able to supply the school with relevant information available at the central office.

External evaluation can also provide the school with information on national standards or benchmarks or comparative data from other schools (e.g. student achievements) that might help the school interpret its own data and assess its quality. Local standards might be narrow in their scope or not available at all. For example, a national evaluation that has focused on school violence prevention projects, can provide schools engaged in internal evaluations of school climate with not only national comparative data on mean rates of violence, but also with a more comprehensive definition of school violence and its measurement.

3. Legitimizing the Validity of the Internal Evaluation
Although internal evaluation should exist in its own right, it is always suspected of being biased and subjective. External evaluation can be as biased, but the accusation of subjectiveness tends to be directed more towards internal evaluation. Being blamed for lack of objectivity is not only a threat to the credibility of internal evaluation, it jeopardizes its existence. An evaluation that is not trustworthy might have difficulties in obtaining the resources necessary
External evaluation can help legitimize internal evaluation but not by providing a formal "stamp of approval" but by respecting its existence as an important ingredient of school evaluation. External evaluation can engage in a dialogue with internal evaluation on equal grounds and with mutual respect, sharing the common cause of gaining a better understanding of the school and its problems. External evaluators don’t have to agree with all the findings of the internal evaluation, but they have to perceive them as good enough to argue with.

**How Can External Evaluation Benefit from Internal Evaluation?**

In many ways and not less than how it can contribute to internal evaluation. Internal evaluation can deepen the scope of external evaluation by increasing its awareness of local issues. It can improve the interpretation of external findings by making them more sensitive to local needs. It can also improve the utilization of external evaluation by diminishing resistance to it and increasing evaluation mindedness (Glasman & Nevo, 1988).

1. **Expanding the Scope of External Evaluation:**

External evaluation is too often criticized for its narrow scope and its tendency to focus on commonalities rather the uniqueness. In its attempt to seek comparability and generalization, external evaluation is forced in many cases to address the "lowest common denominator", which might be either trivial or insensitive (or both) to local needs and priorities. Our own experience in evaluating the Thirty Townships Project (Nevo & Friedman, 1999) is a case in point. This project was initiated and implemented by the Israeli Ministry of Education, with the aim of improving local education systems of thirty townships suffering from low levels of student achievement, high dropout rates and low community self-image. Although the main goal of the project was the same for all townships, various intervention agencies implemented the project, together with each township, developing unique intervention programs to meet local needs and in line with the educational perspectives of the intervention agency. The result was thirty different intervention programs. The Ministry of Education commissioned an external evaluation to focus on project outcomes and to be used mainly for accountability. The external evaluation was expected to supply information about the overall impact of the project and thus used an overall set of indicators, which affected its capability of supplying in-depth portrayals of specific interventions. Internal
evaluation could, in this case, extend the scope of the external evaluation by pointing out the relevancy of additional data reflecting the unique character of a particular school and local educational system.

2. Improve Interpretation of Findings
The external evaluation usually has the advantage of allowing to interpret findings regarding the quality of a specific school by comparing them to other schools or national standards and benchmarks. While such interpretations are usually important and legitimate, they might also overlook the local perspective, reflecting the special needs and opportunities of the school. It is the internal evaluation that can add on that local perspective of school and community. The above mentioned ‘Thirty Townships Project’, is a good example of the limited ability of external evaluation to interpret specific findings regarding local perspectives, needs and constraints.

And as societies become more and more diverse and multi-cultural, it is not only important that external evaluation be complemented by locally specific data. It is also important that local perspectives and values be considered in interpreting evaluation findings.

3. Increasing Evaluation Utilization
Schools that have internal evaluation teams, for whom evaluation is part of their pedagogical and administrative life, have a better chance of understanding the meaning of evaluation and its significance in education. Such schools also have a better chance of appreciating the potential usefulness of external evaluation to the lives of their schools rather than opposing it defensively.

Schools that practice evaluation, by establishing internal evaluation teams or other evaluation mechanisms, increase their institutional “evaluation literacy.” They have a better grasp of the significance of educational evaluation and can speak its language. They gain technical skills to understand external evaluation reports and argue with their findings, if necessary, rather than opposing them.

A school that engages in self-evaluation, and collects information on its pedagogical and administrative activities for that purpose, might increase its self-confidence and be less defensive when confronted with negative findings from an external evaluation. Thus, trying to make use of evaluation rather than opposing it.

A school engaged in internal evaluation will tend to develop a commitment to evaluation, which has been found in various evaluation utilization studies (Cousins & Leithwood, 1986; Shulha & Cousins, 1997) as an important factor affecting the use of evaluation.
TERMS OF COEXISTENCE

If internal and external evaluation are both important and can also benefit from each other, then the remaining question is: **How** can they coexist? My work with schools in various educational systems, in Israel as well as in other countries, leads me to the proposition that when there is an external school evaluation, an internal evaluation mechanism has to be established. And when the school has an internal evaluation team, it should also seek some kind of systematic external evaluation. Internal and external evaluation are both needed and can coexist in a constructive way if an appropriate ground can be created for a **dialogue** between the two (Nevo, 1995).

Of course we all love the concept of dialogue and not many people would dare to argue against it. But dialogue cannot flourish on all grounds and in all climates. Some terms of existence have to be provided to create an appropriate habitat for constructive dialogue. These terms of existence are not self-evident and sometimes quite difficult—though not impossible—to achieve. They relate to three aspects of dialogue evaluation: **conception**, **methodology** and **communication**.

On the conceptual level the following terms have to be achieved:

(1) **Evaluation has to be perceived as a means for understanding rather than judgment.** There is no meaningful way to judge the overall quality of a school by one single criterion or a justified combination (Stake et al., 1997), nor is there any real need to do so. Evaluation is concerned with quality and thus should provide judgmental statements based on multiple criteria related to the various aspects of schools and schooling. It should refrain from an overall assessment of students, programs or schools. Thus, evaluators should provide **quality profiles** rather than **composite scores**, which require justification of weights for the various components that would comprise the composite score. Such profiles could enhance understanding to be used sensibly for improvement, selection, certification or accountability in a specific context and under given constraints. Composite scores provided by assessors tend to be perceived as being “objective” and they are frequently misused or abused for social and political purposes. Actually, such overall judgments are unnecessary. If the aim is formative evaluation, they are too general to provide constructive guidance for improvement. For summative evaluation, making an overall assessment is actually like making the decision and that is the responsibility of the decision maker not the evaluator.
Dialogue Evaluation: Combining Internal and External Evaluation

(2) Evaluation (internal and external) has to be modest, **acknowledging its limitations**, as they are dictated by the state of the art of the profession. Some evaluators tend to promise their clients objective descriptions and unbiased assessments, which they hope to obtain by virtue of their powerful evaluation methodology and instrumentation. Such promises are far beyond the capabilities of current evaluation methods, and are usually an expression of exaggerated professional pride, sometimes bordering on arrogance and deceit.

(3) Although I suggest that evaluators – external and internal – refrain from overall judgments, **providing recommendations** should be part of the responsibility of evaluation. The evaluation literature does not universally accept recommendations as part of the responsibility of the evaluator. Scriven (1995) perceives recommendations as being beyond the scope of the evaluator’s knowledge and expertise. Patton (1997), on the other hand, who is strongly oriented towards evaluation utilization, suggests that “well-written, carefully derived recommendations and conclusions can be the magnet that pools all the other elements of an evaluation together into a meaningful whole” (p. 324). I believe that providing sound, specific and practical recommendations is an integral part of evaluation, and as I shall point out later, a first step in sharing the responsibility for coping with the consequences of an evaluation.

On the methodological level dialogue evaluation requires the following terms:

(1) Evaluation should be practised as a **process** and not as a one-shot activity. Evaluation is a complex process trying to understand complex issues, unless you want to deal with trivia. Sometimes we use figures to understand things. Sometimes we create thick descriptions and portrayals to grasp complex realities. It’s a process of presenting findings, analyzing them, discussing them with pertinent audiences, comparing them with other findings, collecting additional information, getting more findings and coping with added complexities.

(2) The interaction between internal and external evaluators should be based on a **two-way flow of information** in a process of mutual learning. The parties involved in the dialogue are not necessarily equal in their authority, but there is symmetry in the assumption that each has something to learn from the other and something to teach the other. To start with nobody knows **everything**, but each party knows **something**: through the dialogue they learn more and more. Each party can take advantage of its unique sources for data gathering, but for a two-way flow of information both parties, the school and the external
evaluator, have to be engaged in some kind of **systematic data collection** activity. Hence, the school needs an internal evaluation team, or some other evaluation mechanism, to be trained in using available **methods and instruments** for systematic data collection and analysis guided by either a **quantitative** or a **qualitative** research paradigm. Qualitative research may seem more attractive to teachers but is often found later to be too complicated to handle and untrustworthy by parents and administrators. On the other hand, quantitative research methods tend sometimes to be used in a simplistic way of "crunching numbers" with trivial data. Thus, school people should be acquainted with qualitative as well as quantitative methods, understanding both strengths and weaknesses of those research paradigms.

(3) Evaluation should focus on **relevant issues and pertinent data**. A dialogue, any dialogue, is a demanding undertaking, requiring from its participants openness, self-confidence, and a great deal of energy. People tend to avoid dialogue if it is not related to issues that are really important to them. If an external evaluation deals mainly with trivial issues, or is preoccupied with methodological sophistication rather than with issues of substance and important information, school people might lose interest in the evaluation and stop participating in any dialogue. Delineating major issues and relevant information should be an important component of school evaluation as well as any other evaluation.

**On the communication level**, three more terms are necessary:

(1) There must be **mutual respect and trust** between the parties. Both have to believe that each has a genuine interest in understanding what is at stake and can make a significant contribution to such an understanding. An educational system that follows a bureaucratic conception of teaching, defining the role of its teachers as implementers of a curriculum prescribed to them by administrators and experts, doesn’t trust teachers to understand teaching, nor does it trust them to assess its quality. Such a system will probably seek an accountability program based on external evaluation, or employ national supervisors whose job it is to evaluate and supervise teachers’ work rather than improve it. An educational system with a more professional conception of teaching, will expect its teachers to plan, design, conduct and evaluate their work, strive to achieve identified goals and meet defined standards of excellence. Such a system will perceive teachers as professionals, who are interested in understanding the problems of their profession, maintaining its standards and assuring its quality. In an educational system like that, teacher
evaluation can be based on a dialogue between external evaluation and internal self-evaluation of teachers.

Professional evaluators tend sometimes to be arrogant, rationalizing their arrogance by an ideology of isolation, which urges that to preserve their objectivity they have to detach themselves as much as possible from clients and evaluatees to avoid bias and cooption. Such evaluators think they know everything. For them a dialogue with amateur – non-professional – non-objective school evaluators would not only compromise their objectivity, it would also be a waste of time. For a dialogue, professional external evaluators must be modest and respectful to teachers who serve on internal school evaluation teams. Teachers should be respected as professionals and equal partners even though they are not professional evaluators.

(2) Evaluation has to be fair to both parties involved in the dialogue in several ways. First of all, propriety standards have to be observed assuring that evaluation is conducted legally, ethically and with due regard to the welfare of those associated with the evaluation, and those who may be affected by its consequences. Teachers who believe that they are being unfairly evaluated by the principal or by the parents cannot be expected to participate in a sincere dialogue regarding such evaluation.

It must also be clear to both sides what the purpose of evaluation is, what are its expected benefits, what will be its price and who will have to shoulder the costs. An evaluator commissioned by a school to assess its mathematics program for the designated purpose of improvement and innovation, who later discovers that he has been trapped into a power struggle between the principal and some of the teachers, and that there is no intention on the part of the school to improve its math program, cannot be expected to be a good partner for a dialogue.

It is also unrealistic to expect a school to maintain a dialogue with an external evaluation, even one that the school has joined voluntarily, if at a certain point the school feels that it is being exploited for purposes that clash with its needs. This can happen when a school believes that the evaluation mainly serves the research interests of the evaluator/researcher or the information needs of the educational system.

(3) Both parties should take some responsibility for handling the consequences of the evaluation. If a dialogue is to be developed between external evaluation and internal evaluation, there must be a sense of joint responsibility for the consequences of evaluation. If a national or regional authority conducts the external evaluation of a school, and we want to move into a dialogue between the school and that authority regarding the quality of the school, the school cannot be left alone to bear the consequences of such
evaluation. The external evaluation agency must commit itself to some kind of partnership in developing solutions to the problems that will be revealed by the evaluation even when the authority and expertise it has to solve such problems is limited. Professional evaluators serving on or for national or regional evaluation authorities might not be qualified to solve the problems that they uncover. But even they should see it as their responsibility to provide useful and practical recommendations and help the school at least to find the way to those who have the necessary skills and resources. On the other hand, some external evaluators, such as school inspectors in many European countries, have a clear definition of authority and responsibility not only to evaluate their schools but also to facilitate their overall functioning. They should definitely commit themselves to fully share the responsibility to bear the consequences of the evaluation. There is little motivation for a school to move into a serious dialogue for school improvement, if the burden of improvement lies only with the school. Leaving the school alone with its responsibility to solve the problems revealed by the evaluation, not only decreases the chances to find appropriate solutions, it also implies that the school is to be blamed for those problems. Accusations usually arouse defensiveness and counter-accusations rather than stimulating problem solving and constructive action. Unfortunately, demands for accountability, and the use of parental choice as a remedy for school problems, are examples of the tendency of educational systems not to share with their schools the responsibility of bearing the consequences of school evaluation.

**CONCLUSION**

In many educational systems **everybody seems to hate external evaluation while nobody trusts internal evaluation**. I have contended in this article that both types of evaluation are needed as they both have important roles in the life of schools, teachers and educational systems. I have also suggested some conditions that have to be met to establish a constructive dialogue between internal and external evaluation as a basis for their coexistence.

I would like to conclude by suggesting that those of us who are proponents of external evaluation should find ways to empower schools and teachers to participate as equal partners in the evaluation process and make use of it. And those of us who believe in internal evaluation as a means for school autonomy and teacher professionalization must admit the legitimacy of accountability and the right of the public to know. They, in their turn, should seek external evaluation as a partner for dialogue rather than an object for rejection.
Educational officials ask me sometimes: "We believe in external evaluation and we are spending a lot of money on implementing our national school evaluation system. Why should we waste money on internal evaluation which nobody trusts anyhow?" And my answer is: "If you can overcome the resistance to external evaluation and if you think that it is also useful and being used, don't waste a penny on internal evaluation. But deep in your heart you know that this is not the case."

And school teachers often ask me: "We know that you are a strong believer in teacher professionalization, in school autonomy and in self evaluation and reflection. Then why do you agree with external evaluation?" And my answer is: "If you won't respect the authority and responsibility of the Ministry of Education and the right of parents to know about the schools their children go to, don't expect them to respect your right for autonomy and reflection and trust your judgement as a professional teacher."

REFERENCES


INTRODUCTION

It has been called the “age of inspection”, the “evaluative state” and the “audit society.” Whatever term one prefers, there can be little doubt that something systematic has occurred since 1971. In every area of social and economic life, there is more formalized checking, assessment, scrutiny, verification and evaluation (THES, October 18, 1996).

Schools which evaluate themselves against self-generated critical standards and document their progress demonstrate both their own professional development and public accountability (Simons, 1995).

The first quotation above taken from an article titled, “I audit, therefore I am”, by Michael Power, Professor of Accounting, London School of Economics, signals the main thrust of evaluation that is prevalent in public policy today – that of external accounting or surveillance of social systems.

The second quotation indicates a quite different perspective, that the onus for evaluating and monitoring one’s performance and practice lies with institutions and professions themselves. This is not to say that no external accounting is necessary. Rather that the emphasis should be on evaluation conducted by those within the institution.

In this chapter my main task is to focus on school self-evaluation in a democracy. I first outline the case for schools evaluating themselves and present the underlying philosophy and origins of a democratic case study approach to
school self-evaluation. Secondly I explore purposes and process and suggest
how schools might undertake this task, simply and efficiently taking into account
the teaching load and responsibilities they carry. Thirdly I outline the philosophy
of a training programme designed to support schools in this process. Finally,
I examine how the process may be strengthened to include other citizens
and to justify the integrity of school self-evaluation to the public, yet maintaining
the emphasis on the school evaluating itself. I start by locating the place of
school self-evaluation in the governance of a democratic education system.

JUSTIFICATION FOR SCHOOL SELF-EVALUATION
IN A DEMOCRACY

That evaluation of schools is needed in a democratic society is not in question.
What is in question is the kind of evaluation that is needed for what purpose.
Currently in many Western democracies, student performance testing and school
inspection are held up as yardsticks of how well schools are doing and provide
the basis for judgments of a school’s worth and school choice. Such judgments
are narrow in the extreme and fail to do justice to the complex learning and
teaching that takes place in schools and the many and varied achievements of
students. There is nothing very democratic about a system of evaluation that
does not value diversity of achievement, autonomy of individuals and co-
operative attempts by schools to promote learning conditions that acknowledge
such differences and potential. Neither, is there anything very democratic, in
terms of promoting social equality, in publishing exam results in league tables
deliberately to distinguish the best from the worst. One could argue that putting
such results in the public domain is a democratic act. Let the citizens decide.
But this is hardly an adequate response. The data on which the citizens are
being asked to decide is one-dimensional, limited to only the outcomes of
schooling in particular subjects. So much more is part of the educational process
and should be taken into account in a fair assessment of school’s worth. As
Stake (1978) has noted:

The worth of a program is seldom indicated by the achievement of students. That is partly
true because our measurement instruments are narrow and crude. They indicate only a small
part of the impact of a lesson or program . . . . It is also partly true because the worth of a
program is dependent on its relevance to other courses and programs, on its effect on
teachers’ morale, on its compatibility with community ideology, etc.

Clearly, our database for evaluative judgement needs to be broader if it is to
be fair.
**School Self-Evaluation in a Democracy**

*Evaluation for Different Purposes*

In eschewing league tables of results as the single criterion of a school’s worth, I am not saying that some form of external testing of students and inspection of schools is not needed in a democratic society. House (1992) has made the point in the context of the governance of education that different forms of evaluation are legitimate and necessary for political and professional reasons. Drawing on Lundgren's models for the governance of education, he identifies four types of idealized evaluation procedures on a central-local axis and a professional-political axis. The kind of evaluation I am advocating in this paper falls within the local-political cell of the matrix. In other words, schools should be accountable to the public for the education they provide, primarily, though not exclusively, I would add, at the local level. In the local professional cell of the matrix are teachers-as-researchers, i.e. teachers researching their own classroom practice. In the central/political cell there would be external testing of pupil performance while large-scale external evaluations of the state of the art of education or evaluations promoted by national professional groups would be located in the central/professional cell.

These ideal evaluation procedures are based on two assumptions. The first is that the primary causal factors of classroom learning are under the direct control of teachers who essentially determine classroom processes whatever external mandates are imposed. The second is that other groups – politicians, the parents, and the local community – have a right to public knowledge about the education schools provide (House, 1992).

The critical political question is what should be known and who should provide this information. Central governments and/or local or state governments need broad indicators of the state of the nation in particular subject areas, but they do not need to know everything that goes on in every school. (A light-sampling testing programme of certain subjects, for instance, could meet this political information need). Schools need a degree of autonomy, free from outside scrutiny, to experiment, and take those creative risks that fuel good education and motivate teachers and students to improve.

These are different evaluative purposes and require different processes. The problem arises when they are confounded. Currently, at least in the U.K., we have a situation where the information needs and purposes of different kinds of evaluation are not recognised and acknowledged. One kind of information (league tables of performance achievement) is presumed to serve the other (i.e. to be an indicator of a good school) when league tables of exam results are only one factor to be taken into account in determining what is a ‘good’ school. Whether they are a measure of the quality of education is very debatable indeed.
The central political need for information on standards and performance achievement overwhelmingly dominates as an evaluative measure of education.

Such a focus on short-term political needs is not in the best interests of education. Much lip service is given to the adage that ‘teachers are our best resources’, but little investment is provided for the kind of professional development programmes that would make them an on-going creative resource. In the next section I outline a case for school self-evaluation that is educationally motivated yet politically responsive. Its basic aim is to improve education in the long term, but it takes account of the fact that some political response is needed in the short term. While the focus is on teachers and schools evaluating their own policies and practices, other educational groups in our society have a role to play in supporting the conduct of evaluation by schools themselves. Sharing responsibility is part of the democratic process. I return to this issue in the final section of this paper.

THE CASE FOR SCHOOL SELF-EVALUATION

The case for school self-evaluation outlined in this chapter is based on the following premises:

that teachers are in the best position to evaluate curriculum and learning changes at the point of need, responsive to students they know well and whose progress they are continuously assessing;

that the quality of education can best be improved by supporting and strengthening the collective professional development of teachers, schools and school districts;

that this is best achieved through the creation of a collaborative working culture in which teachers, schools and schools’ districts can safely, critically and publicly evaluate their work and the conditions in which they work;

that schools and districts which implement such a process of on-going monitoring and self-evaluation of their policies and practices and open this to public scrutiny, demonstrate in so doing, their accountability to students, parents, the profession and the wider public.

Explicit in the above premises is a commitment to quality of education through respecting and supporting the autonomy of professionals within a shared collaborative culture. This culture operates at three levels; between professionals within the school; between professional groups – teachers, local education authority advisers for instance – and between these and their major constituencies, parents, students and the wider public. This is underpinned by shared responsibility and mutual accountability for the quality of educational opportunities provided by the school and supported by the community.
Origins of this Approach

The case for school self-evaluation just outlined has its origins in the evaluation methodology that accompanied the curriculum development movement of the late sixties and seventies in the United Kingdom and the analysis generated through that of the failure of external reform efforts to impact upon schools. Two aspects of the evaluation methodology are particularly relevant: the case study approach and the political democratic model.

The reasons for the evolution of the case study approach in the curriculum reform context has been documented in many places (see, for instance, MacDonald & Walker, 1975; Simons, 1987, 1998) that I feel no need to reproduce them here. However the essential features of the approach are worth noting for it is these that make the approach also applicable to the self-evaluation of institutions. First, context became central to an understanding of the curriculum in action. Secondly, participants were acknowledged as active agents in the process of the evaluation. Thirdly, relationships between people and the politics and management of the institution had profound effects on implementation. Fourthly the culture of each school and different regions of the country led to quite different effects in practice making it difficult to generalise.

Case studies of individual schools not only generated an understanding of the complexity of projects in action in the uniqueness of the culture of the school and region but also allowed changes over time to be documented in the context in which they occurred. From these early beginnings case study evaluation has burgeoned and it is now a widely accepted and legitimated approach in the study of complex social settings. (See, for example, Stake, 1995; Simons, 1980; Merriam, 1985.) While the case study approach was generated in the context of external evaluation, it is easy to see its adaptation to internal evaluation. Its focus on the institution,1 its context and uniqueness and the processes, methods and styles of accessible reporting that it promoted, are within the scope of all professionals whether external or internal to the institution. The scale and purpose of the exercise may differ, and possibly the initial confidence of staff but the techniques and processes are applicable, whether working from within or without.2

The second development in evaluation methodology that occurred around the same time was the acknowledgment of the political dimension in evaluation and in particular the evolution of the democratic model of evaluation. (See, for example, MacDonald, 1974; Simons, 1987, 1998.) This was a counter response to prevailing trends in evaluation methodology at that time, which failed to acknowledge sufficiently the political and ethical implications for people in the conduct and dissemination of evaluation. It was also a response to the increasing centralization of power and control in education.
Briefly stated, the concept of democratic evaluation has as its central aspiration how to find an appropriate balance between the public's right to know and an individual's right to privacy in the conduct and dissemination of evaluation. The concept derives from the rhetoric of liberal democracy, a rhetoric that is broadly acceptable to those exercising delegated power. From this rhetoric is derived a set of power-equalizing procedures that cut into the customary relationships embedded in organizations holding participants accountable to criteria endorsed by them. Central precepts in the procedures are confidentiality, negotiation and accessibility. Such procedures cannot, of course, change the power relationships, but what they can do is to accord equal treatment to individuals and ideas, establish a flow of information that is independent of hierarchical or powerful interests, and ensure that no one group or person has the power of veto. In such a context all relevant perspectives can be represented, information fairly and equitably exchanged, and deliberation encouraged.

In the context of school self-evaluation, one might think at first that the central precepts of democratic evaluation are not so applicable. Is it possible to maintain confidentiality, for instance, in a context where everybody knows each other? However it is surprising within schools how little is known about colleagues and how little colleagues wish their ideas, values and practices to become public. Schools in many respects are private places. In evaluation values and perspectives that hitherto may have been private become part of the public agenda. The procedures of confidentiality and negotiation are still relevant however to gain confidence and secure honest data. Knowing there is a process of negotiation will reassure participants that they have some control over data of a personal/private nature. If such data is relevant to the evaluation this needs to be negotiated to become public (even within the school). Other data which, it can be argued, is part of a teacher's and schools' professional life, is already in the public arena. Confidentiality is also important in gathering data from students to protect their identities and to reassure them that they will not be exposed to comment or unfair censure from teachers for what they may have said in interviews. Negotiation serves the further function of establishing when findings from the evaluation are released, to whom and in what form.

Theory of Change Underpinning School Development

The concept of school self-evaluation advocated in this paper is grounded in a theory of school development that evolved from external evaluation of centrally developed curricula taking account of the failure of these central initiatives to transform schools.
I have no space here to detail all the components of this theory but it acknowledges, among other things, the school as the basic unit of change; the professional autonomy and direct accountability of teachers and the need for a strategy of change that is collaborative and developmental. This not only gives teachers ownership over the change process but also acknowledges the inherent curriculum theorising and teacher development that teachers do. Bringing to conscious awareness an understanding of the implicit theories and values that guide and sustain teaching and curriculum practice within the school, self-evaluation can directly and developmentally enhance the educative process. Finally this theory of school development acknowledges that change is a professional community activity. Teachers, schools, local education authorities and parents need to work together to support and facilitate change. The recent growth of consortia arrangements between schools, local education authorities and higher education institutions to promote teacher research also acknowledges the professional community support that is needed for change. For further elaboration of this theory of change see Simons (1998).

SCHOOL SELF-EVALUATION: PURPOSES AND PROCESS

That, then, is the case for school self-evaluation. What does the process look like in practice? I start with a broadly acceptable definition. School self-evaluation is a process of conceiving, collecting and communicating information and evidence for three distinct purposes:

- informing decision making within the school, i.e. to facilitate a process of school development;
- ascribing value to a policy within the school or to the school itself;
- establishing public confidence in the school.

There are several aspects of this definition that warrant underlining. The first is the emphasis on evidence. This is important to distinguish the process from other processes that offer judgments on schools without a research or evaluation base. It also signals that unsubstantiated assertion and opinion or non-triangulated data is insufficient as a basis for judgment and action.

The second is value. This is a reminder that evaluation is essentially about establishing the value or worth of something and that ascribing value is a process undertaken by people. It is not embedded in evaluation instruments such as tests or questionnaires.

The third is communication. While this may seem an obvious point to make, it is surprising how often little thought is given to communicating the results
of evaluations in an effective way. However if results of evaluation are to feed into policy making within the school and establish confidence with constituencies outside the school, attention needs to be given as to how best to do this. The process of school self-evaluation I am outlining here is an on-going process that has implications for development and action. It needs to be accompanied by an inbuilt strategy for communicating and discussing the findings, first among staff and secondly, with parents and the wider community.

Finally, I wish to emphasize the purpose. School self-evaluation, like any evaluation, can be undertaken for a variety of purposes, not all of them compatible. Macbeath (1999) outlines six different purposes – political, accountability, professional development, organizational development, the improvement of teaching, and the improvement of learning. Others prioritise external accountability (e.g. OFSTED, 1998) or combinations of the above (e.g. Nevo, 1995; Simons 1998). What is important is to be clear which process or set of data serves which purpose and not conflate one with the other. School self-evaluation for improvement of teaching, for example, requires a direct focus on teaching and fulfils a professional development purpose but not necessarily a political/accountability purpose as House (1992) has indicated. Arguing a dual purpose or establishing interrelationships between different purposes is a different point. In my own work (see Simons, 1985) I have strongly linked school self-evaluation for professional development with institutional development to acknowledge the change in value structure that is often needed to support the process of self-evaluation within institutions.

Self-Accountable Professionalism

My own position also links professional development with accountability, but accountability in a particular sense. School self-evaluation has often been introduced into schools or local districts in response to external pressures to gain more information (some would say control) about schools or more accountability from them. However accountability that stems from within, as already intimated, has intrinsic value as well in strengthening the confidence of schools to meet their educational responsibilities. Paradoxically this is a stronger form of accountability than that which is externally sought.

Self-accountable professionals – the term I have adopted to refer to those who take this stance – not only reflect upon their practice to improve the internal working of the school. They evaluate what they do against self-generated critical standards, they research shortfalls in provision and performance, they respond to changes of context and clientele, they experiment, and they evaluate and develop new programmes to solve identified problems. This does not mean that
School Self-Evaluation in a Democracy

their work is insular or confined to the classroom or school. Self-accountable professionals also collaborate and engage in persuasive negotiation with the constituencies whose support they need, and make the findings of their deliberations accessible to legitimate audiences outside the school. (For other definitions of what it means to be professional see House & Lapan, 1988; Hoyle, 1975; Wise et al., 1984). It is a responsive community-based professionalism, whether that community is a community of teachers within a school or between schools, or between schools and local districts or between schools, their parents and the broader community. Such an approach provides the basis for the recognition that Reid (1987, p. 14) believes professionals need:

...we should remind ourselves of the importance for professional practice of the support of publics outside the professions and beyond institutions... the theory on which professions act gets its essential meaning and significance from the outside world, not simply from the consensus of those within a profession. It is therefore essential that they address themselves to these wider publics to implant in their minds the image of the profession which those within it would like to own.

Forms and Processes

There are many different forms of school self-evaluation and different motivations for introducing them. These include checklist approaches such as those promoted by local education authorities in the mid-seventies in the U.K.; organizational development, such as those promoted by the GRIDs Project (see, for instance, McMahon et al. (1984), curriculum and learning focused approaches introduced by individual schools (see, for instance, Mitchell, 1984), policy-issue based approaches that are criteria-led (such as the one I outline below) and broad frameworks such as the one recently outlined by MacBeath (1999). I do not have time to outline and/or critique these different approaches here, though some critique may be found in Simons (1987, 1992).

However, three major deficiencies of early forms of school self-evaluation, such as the checklist approach, are worth noting. The first was the failure to incorporate a process of self-evaluation into the on-going structure of the school. Self-evaluation was an add-on, not an integral part of the essential functioning of the institution. The second, sometimes true of the organizational and issue approach as well, was the failure to focus sufficiently on teaching and learning issues. Where choice of topics to evaluate was offered in such approaches, there was a tendency to choose institutional issues, not classroom-focused issues. While broader institutional issues are important to evaluate for understanding and generating an appropriate climate for school self-evaluation, a focus on these issues can also be interpreted as a safe option compared with the more
potentially threatening process of looking at teaching and learning. The third was the failure to embrace a strategy of change that allowed for dialogue and conflict over values to be expressed.

**Issue-Focused Approach**

One of the major aspirations in advocating a process of evaluation that schools can undertake themselves is that it be feasible to conduct. This means that it has to be economical in terms of time; within the competence of staff; and supported by the whole school with time and resource support set aside for the purpose. For the process to be valued and to feed into school development, it needs to be on-going, built into the structure and functioning of the whole school. The purpose needs to be clear to all and the focus limited to a study of one or two issues at a time within a framework of priorities chosen by the staff and related to school development plans where they exist.

Taking account of some of the deficiencies of earlier models, I suggest the focus is primarily on curricular policies and practice – policies which affect all students and to which all teachers in the schools should be committed. Issues of teacher development and pupil performance can be considered within the context of these whole school policies.

The process itself need not start with collecting new data. Much data already exists in the school that can be used to raise questions and focus an in-depth study. If an on-going monitoring system has been set up, this will provide a backdrop for other issues studied in depth. Where data has to be collected, it is useful to use ‘low technology’ techniques – such as interviewing, observing, analysis of documents, critical incidents, – which are familiar to teachers in the context of teaching and can be adapted for the purpose of evaluation. It is also useful to remember to engage the skills of members of staff in the art, mathematics, and information technology departments to name a few. They can offer specialized help with analysis and presentation as well as be part of an evaluation team.

Finally it is important to underpin the process by the adoption of a collaborative, participatory ethic involving as many in the school as possible, (though not all at the same time) and to timetable forums for the discussion and implementation of findings.

In summary, school self-evaluation is a process whereby small teams of teachers (and possibly students, parents and governors) evaluate, on behalf of the school, an aspect of the curriculum or a cross-curricular issue in depth. The team, in consultation with the school, would determine the criteria for evaluation relevant to the chosen topic, gain evidence from a range of personnel
and sources, analyse the data and communicate the findings in written and oral form to the forum within the school to discuss implications for development. They might also provide a brief case study account to audiences beyond the school. The aspiration to be self-accountable professionals is met by seriously undertaking the process. Giving an account to relevant constituencies outside the schools strengthens this commitment.

**Implementation and Support**

Though this approach to school self-evaluation was generated over twenty-five years ago, the democratic values that underpin it are still prevalent today. In a context of highly centralised government edicts, evaluation on democratic lines, it can be argued, is needed now more than ever before as a counter-balance to the managerial forms of evaluation that dominate education. As Handscomb, (1995, p. 10) has written commenting upon some of the negative aspects of local governance of schools in the United Kingdom:

> The Government is right in linking quality education with school responsiveness, but this in itself is dependent on a shared, democratic purpose at community level.

It is within this ‘democratic sense of purpose at the community level’, that school self-evaluation has thrived over the past twenty-five years in certain countries at certain periods of time where the conditions were propitious for such development. To thrive at a local community level often needs political support at central government level. In other words there needs to be an acceptance of the value of the enterprise by the governing powers of the state if the process is to have a genuine chance of being implemented effectively to generate changes in schools. There is little point in telling schools they should self-evaluate if the philosophy underlying that process is not strongly endorsed by central government or it is undermined by government policies pursuing contrary agendas. The commitment has to be genuine and preferably underpinned by training and professional support.

**Training for School Self-Evaluation**

Training in school self-evaluation is not a common occurrence. It is another task schools are asked to undertake without additional support. Yet those who have been through the process suggest training is needed in a variety of evaluation skills. While it is true, and indeed part of the case for school self-evaluation (as indicated above), that teachers can build upon methods and processes they constantly employ in teaching, such as observation, questioning,
shadowing, and critical incident analysis, the use of these methods in evaluation is different – devoted to a collection of evidence for analysis rather than direct teaching. There are also other evaluation skills to be learned connected with, analysis and presentation of findings and the sharing and dissemination of knowledge to other groups who may or may not want knowledge of each other. In other words, in evaluation, political and interpersonal skills become as important as the technical skills of gathering and analysing data. Based on research into how schools handled implementation of the process of school self-evaluation and the research on change previously mentioned, I generated a training programme designed to support teachers and school districts conduct case study self-evaluations in their workplace contexts. The programme had the dual aim of serving the professional development needs of teachers and providing an account to local communities.

DESIGN OF TRAINING PROGRAMME

The overall purpose of this training programme is to enhance student learning and potential within schools by a conscious examination of particular policies and their effects. The programme takes place over four months and in three main phases. First, it involves concept and methods training of six days spread over six weeks (to incorporate working with data from case study schools). Secondly, and concurrently, teams from individual schools and local districts conduct a case study evaluation of a policy issue in their school or district. Thirdly, each team presents their written case study evaluation to the other teams and presents a critique of one of the other team’s evaluation.

The training process itself is also predicated on sharing and peer critique between levels (primary, secondary, administration) and within levels. The case study is limited to ten pages. Creativity in design and presentation is encouraged. The philosophy underlying the design of the programme is given below. The actual programme and results from the evaluation of it are reported in Simons (1995).

Assumptions Underlying School Self-Evaluation Training Programme

(1) School self-evaluation should build upon the informal evaluation processes schools already undertake as part of the normal functioning of the school. Often these are non-systematic and private to the classroom or school. In school self-evaluation, these become formal, systematic and public and time may be needed to adjust to the process. While there is currently a great deal of external pressure on schools to be publicly accountable on practically
every dimension of a school’s activity, as in OFSTED\textsuperscript{5} inspection for instance, such a stance does not give teachers and schools the scope for taking charge of their own developmental process. The emphasis is on demonstrating performance for external clientele rather than improving education from within. Too much dominance on an external process also detracts from the more immediate task of feeding results of evaluation creatively into their own development. This is why I advocate school self-evaluation as an ongoing, continuous part of school development. The task for a self-accounting profession is to gradually shift from the informal to the formal in evaluating and documenting the work of the school.

(2) The essence of producing a good quality monitoring, review and evaluation system is teamwork; developing a collective professional responsibility and creating a collaborative culture to support the process. The design of the training programme therefore includes teams of teachers from the same school and teams of education officers from the same districts working together; and sharing the process and critique of their work with other institutions at the same level and between levels (primary, secondary, administrative office).

(3) School self-evaluation needs support at different levels in the education system. Personnel working at these different levels in the system need to share assumptions about the concept, process and expectations of school self-evaluation so that all are working to the same understanding. The design therefore emphasizes the administrative office also conducting a case study of a policy issue: to mirror the process; to understand the support needed for implementation and to consider the training needs so they can contribute to future training for other schools and districts.

(4) If school self-evaluation is to facilitate development and become part of an ongoing monitoring exercise, it needs to be economical (within the timescales in which teachers and administrators work), feasible (within the competence of staff) and useful (meaningful and accessible to other staff and audiences to whom the results are disseminated).

(5) School self-evaluation needs to be premised on a growth model not a deficit model of change if it is to have any chance of improving educational practice.

(6) Evaluation which seeks to improve educational practice requires in-depth analysis. For this a clear focus is necessary. The design therefore focuses on schools and districts evaluating one policy issue in depth at a time in the context of the school and district’s development plan. The background is the school or school district development plan; the foreground is the priority issue chosen for discussion. Large-scale, comprehensive reviews
may be useful to provide 'state of the art' accounts. However their very comprehensiveness can lead to superficiality, militate against direct feedback into decision making, take over from the essential work of teaching and learning for a period and leave all involved 'heaving a sigh of relief' until the next comprehensive review comes round.

(7) Evaluation itself should be educative for those conducting the evaluation, those participating and those to whom findings are disseminated, i.e. it should add to the self-knowledge of individuals and the self-knowledge of groups. It should not deskill those it seeks to develop.

(8) External mandates, while acting as a catalyst for action in some instances, do not provide the most appropriate motivation for school development. Intrinsic motivation has more lasting staying power and is likely to be more enjoyable. Schools, teachers, administrators will become their own best critics if they have some control over the evaluation process, over the choice of issues to evaluate, the methods and procedures to be employed and the audience to whom results will be disseminated.

(9) Far from being threatening as is sometimes feared, constructive critique is an important learning process built into the design of the programme at three levels: critique of each other's plans, the methods and procedures to be adopted and the final case study analysis and report.

(10) A school self-evaluation-monitoring programme that aspires to become a developmental process needs to involve the whole school at some stage in the process and to be underpinned by an effective strategy of change. For me this is one committed to collaborative values based upon shared goals, shared ownership, shared skills, shared responsibility, shared process, shared analysis, shared findings and action.

STRENGTHENING THE SCHOOL SELF-EVALUATION PROCESS

School self evaluation as the name suggests is a process conducted by the school itself. This leaves open the question as to who constitutes the school. Is it the teachers and students only? Or ancillary workers within the school and parents and governors who support the school? All of these I suggest as they each, in their different ways, contribute to the life and culture of the school. However it will be teachers in the main who generate and sustain a process of school self-evaluation, which is why I have focused on their individual and collective professional development as a major aim of the process.

It is also useful to recall how and why the process of school self-evaluation first developed. In practice, the initial focus was on teachers to address the
deficiencies in early reform efforts that failed to engage teachers in the process of reform. Later the focus shifted to the institution to strengthen the location in which change takes place. It is useful to remind ourselves in this context that whatever national initiatives and guidelines exist (and schools, of course, have to work within these), education actually takes place locally, in local classrooms, local schools and local contexts with particular teachers and particular students.

A third historical point is that engagement in the process of school self-evaluation was innovative for many teachers and innovation, as we know from previous reform efforts, is a challenging activity that takes time. Teachers needed this time to become familiar with the idea and to gain the skills and confidence to conduct the evaluation, document the results and share the findings. Having said this, it is also true to say that once teachers gained this confidence, involvement of students, parents and outsiders in the evaluation process was seen as a welcome step to realize the democratic aspirations of the process. Given the pace of innovation and change in the eighties and nineties and the public gaze schools are under, teachers are perhaps more confident these days in undertaking new endeavours, though there are always likely to be risks associated with making the private public.

The issue in the current climate is how to strengthen and validate the process of school self-evaluation as a major vehicle for school development and community judgement of schools in the face of competing, narrow indicators that promise to tell what a school is worth.

When producing self-evaluations for external audiences, it is extremely useful to include other participants from the broad constituency of who constitutes a school to contribute their perceptions on the issue, to triangulate the data and strengthen the validity of the conclusions. There may be some issues, related to teaching and learning perhaps, where teachers and students are the main protagonists in the evaluation team. Evaluation of other issues, such as equal opportunities or multi-racial policies, may benefit from having a wider range of constituencies, parents or governors, for example in the evaluation team.

It is often suggested that a means of strengthening the process of school self-evaluation is to have external evaluation and internal evaluation working in consort in some way. Alvik (1997) indicates that these may be parallel, sequential or cooperative. Nevo (1995) describes a systematic sequential process in Israel where external evaluation takes place once the school is confident with its internal evaluation process. This process is similar to the one I have described above built into the school organisation as an integral part of the functioning of the school, preceded by training and support from an external tutor in that training. In the model I have outlined above I do not assume that self-evaluation is a precursor to
external evaluation – the major purpose is to enhance school development. However schools that have such an on-going self-evaluation process are in a strong position to respond to external evaluation should they be required to do so. There are several other ways in which the process might be strengthened.

First, more attention needs to be paid to changing the organisation of the school to create appropriate conditions for school self-evaluation to take place. Such a change essentially rests upon four factors. The first is having a united purpose that embodies the process of school self-evaluation as an aspiration – it is important that the whole school endorses the activity. The second is developing long term strategies to effect a shift in the value structure of schools if it is not already based upon democratic principles. (For an indication of how this might be achieved, see Simons, 1987.) The third, and more immediate, is finding time within the school day, built into the timetable for teachers to work together on evaluating identified issues. The fourth factor is resource support from local administrations to underpin the work in the schools. This support may be in terms of commitment to the philosophy of the process, to releasing personnel on a consultant basis and to providing finance for relevant materials, papers and exchange visits to other schools undertaking a similar process.

The second development which would strengthen the process is the pairing of schools adopting a process of self-evaluation underpinned by the following strategies: nominating a member of each team as a consultant to the team in the other school and vice versa; and team meetings in one of the schools to extend their understanding of the process in practice to other schools in their district. These strategies are based upon the principles of the power of critique as a strategy of learning and support for schools undertaking the process and the commitment, especially if the process has been supported from the resources suggested above, to extend the work to other teachers and schools.

Finally, I think that the process of teacher development and school accountability could both be greatly enhanced if consortia of schools in the same district focused on evaluating the same issue at one time in order to build up a cumulative evaluation knowledge base that might inform policy. One of the criticisms often cited of school self-evaluation is that it only reflects one school’s concerns. Such a focus would redress this criticism.

The wider aspiration for establishing a cumulative, common knowledge base to inform policy may take some time to reach. However a local or indeed central administration, which was prepared to put resources into supporting a consortia of schools working on a process of school self-evaluation to this end, would, I believe, find that both the professionalism of schools would be greatly enhanced and the knowledge about how schools are developing would be fed much earlier into policy debate.
NOTES

1. Not all case studies focus on the institution. As Stake (1995) has pointed out, it is possible to have a case study of a classroom, institution, policy and education system. For the purpose of this paper I am primarily writing about case studies of institutions, conducted in this instance by the schools themselves.

2. Not all evaluators would agree that teachers can conduct professional evaluation. Scriven (1967, p. 53), for instance, is quite clear that participant teachers are too unskilled to do a professional job. Only specialist evaluators will do.

3. In a previous study (see Simons, 1987, p. 128), I found that teachers were more concerned about sharing their accounts and value positions with colleagues within the school than they were about publication beyond the institution.

4. In the U.K. recently, OFSTED have pronounced that schools should undertake school self-evaluation as a precursor to OFSTED external inspection. While the rhetoric adopted to advocate this initiative is similar in some respects to that suggested in this paper for engaging in school self-evaluation, the origin, motivation and emphasis of the OFSTED initiative is quite different. However it is to be hoped that schools will use the process to good account to develop their own curriculum and institution.

5. OFSTED (1998) of course, as previously mentioned, have recently suggested a self-evaluation process for schools. But here I am referring to their external inspection process.

REFERENCES


SCHOOL SELF-EVALUATION: ORIGINS, DEFINITION, APPROACHES, METHODS AND IMPLEMENTATION

Jaap Scheerens

INTRODUCTION

The origins of school self-evaluation are situated in the restructuring of educational systems, developments in evaluation methodology and technology and in substantive social scientific research and theory. School evaluation can be defined on the basis of an analysis of the evaluation concept and the structural context of the school within educational systems. Since evaluation is closely related to the issue of quality in education and school evaluation approaches are close to all kinds of systems for "quality care" and quality control within organizations - some attention has also to be paid to the concept of "quality" of schooling. The core of the chapter consists of an overview of different types and approaches to school evaluation, geared to different conceptions of "quality" in education. A taxonomy, in which various methods of school evaluation are related to the specific information providers and the specific audiences of each, is developed, as well as an overview of qualitative and quantitative methods. In a final section implementation issues and applicability in developing countries are discussed.
The main conclusion of the chapter is that school self-evaluation can be used efficiently as an “add-on” to external types of school evaluation, actual applications are mostly hybrid forms in which basic types are combined or integrated, and that the process of developing school self-evaluations approaches and tools together with schools has an important innovative potential.

1. ORIGINS

The upsurge of school self-evaluation in European countries during the last decade has societal and scientific origins. Decentralization of educational systems, as the official policy in many countries, has evoked increased interest in accountability, responsiveness and self-improvement of schools. Scientific developments matched these trends, on the one hand through a broadening of educational evaluation methodology and on the other hand by conceptualization and research in the field of school effectiveness and school improvement. These contextual developments will be sketched briefly, before zooming in on the core contents of defining and describing the variety of approaches and methods of school self-evaluation. Before a more detailed underpinning of the definition of school self-evaluation is provided, a working definition of school self-evaluation that is implied in this section is that it concerns a type of educational evaluation at school level that is initiated and at least partly controlled by the school itself.

1.1. Functional Decentralization

The concept of functional decentralization is useful for describing developments in the governance of schooling in European countries, because it recognizes options where centralization takes place in some domains, for example the curriculum, and decentralization in others, for example school finance. The validity of this conceptualization of (de)centralization is shown in OECD, 1998 and Van Amelsvoort and Scheerens (1997). Schools taking responsibility for their own assessment and quality control can be seen as an intrinsic part of school-based management, as a consequence of increased local democracy and local control over schools and as a spin-off of administrative accountability requirements. Increased interest in accountability and centralized output control sometimes appears to function as a counterbalance of devolution of autonomy of schools in other domains.
School self-evaluation functions as a spin-off of assessment procedures that are externally induced for accountability purposes when assessment results are fed back to schools and discussed among staff for school development and school improvement purposes. Such applications can be seen as 'borderline cases' where it can be debated whether such practices are to be seen as genuine school self-evaluation activities. The results of two European Union funded research projects (Van Amelsvoort et al., 1998; Tiana et al., 1999) indicate that hybrid forms of external and internal school evaluation occur frequently.

The second type of combination of school self-evaluation and accountability oriented assessment occurs when schools gather and present additional information which is to present the local stakeholders with a more complete picture of what the school has to offer than pupil achievement scores. Study of parents' criteria in school choice indicate that such additional information concerning, for example, the pedagogical mission of the school and special programs, should be quite relevant (Lam, 1998). Generally speaking such additional information will consist of description and evaluation of input and "process" characteristics of the school.

1.2. Developments in the Methodology and Instrumentation of Educational Evaluation

During a recent symposium on the future of program evaluation in education a broad range of evaluation approaches and strategies were discussed (AERA, 1999). At the end of the meeting the members of the panel (among which Bob Stake, Daniel Stufflebeam and Eva Baker) agreed on a few trends. With respect to the relationship between evaluator and practitioners it was concluded that the traditional distance between both was giving way to a more participatory approach, where the evaluator becomes more of an advisor to the professionals in the situation that is to be evaluated. Objectivity can also be seen as a common interest, particularly when evaluation takes place in a less "high stakes" situation, and can, to some degree, be guaranteed by the methods and instruments that are used. Conclusions were also drawn about types of educational evaluation that were expected to survive into the next century. Assessment programs, indicator systems, and utilization focused evaluation were seen as being among the survivors.

Participation, assessment and monitoring approaches and a keen eye for the way evaluation results are to be used are all important ingredients in school self-evaluation.

The fact that these trends are perceived among educational evaluators indicates that there is likely to be sufficient professional input for instrumentation and capacity-building for school self-evaluation.
As a result of developments in related evaluation sub-domains like educational assessment, school effectiveness research, the use of system-level education indicators and computerized school management information systems and pupil monitoring systems, there is an abundance of instruments available that can be used for school self-evaluation purposes. And there are, of course, also sets of instruments that have been developed for the prime purpose of school self-evaluation. The mere fact that these tools are available provides part of the explanation of the increased scale of application of school self-evaluation in the European context. To some extent the medium appears to carry the message although the context of increased school autonomy and accountability requirements coincides with such a technocratic explanation.

1.3. Substantive Results of School Effectiveness Research and Conceptualization of School Improvement

In reviews of the results of school effectiveness research evaluation practices at school and classroom level feature among the categories that are considered to be positively associated with pupil achievement (cf. Scheerens & Bosker, 1997).

In attempts to ground these empirical research findings in more established theory, like planning theory, public choice theory and organizational images like the one of the learning organization, the centrality of evaluation and feedback mechanisms is evident (Scheerens, 1999).

In the related field of school improvement, that is more design-oriented and makes use of qualitative research methods more frequently than school effectiveness research, school self-evaluation has traditionally been considered as one of the levers to bring about change and innovation. In this context the term “school based review” is most frequently used, for approaches that are primarily based on self-description and self-judgement (Van Velzen et al., 1985).

An evaluation centered approach to school improvement provides an alternative to more proactive, planning approaches which start out with mission statements and objectives. As West and Hopkins (1998) note, such proactive approaches may lead to “laudable vision statements” that schools seem to be unable to reflect in practice. Instead they recommend schools to start the improvement process by evaluating “the quality of experience they currently offer their students”. Scheerens (1999) relates this approach to “retroactive planning” and “result oriented school management”. Such approaches have the advantage of having a starting point that is firmly grounded in the empirical reality of school life, but may be criticized for leading to change that is likely to be incremental rather than synoptic. On the other hand, at least in the Netherlands, experiences
School Self-Evaluation

with more proactive planning in school development have been quite disappoint-
ing (Van der Werf, 1988; Friebel, 1994).

The fact that academics working in the field of school effectiveness and
school improvement in Europe increasingly recognize the importance of
evaluation and monitoring activities at school level is considered as one of the
factors that contributes to the increased application of school self-evaluation.
The upsurge of school self-evaluation as such is seen as the coincidence of the
three contextual factors: functional decentralization and increased school
autonomy in some domains, the availability of methods and tools and the frame
of reference of school effectiveness researchers and of experts on school
improvement.

It is an interesting question why school self evaluation does not to appear to be
an issue in school restructuring in the USA. In overviews of the field like e.g.
Conley, 1997, the term school self-evaluation does not occur, nor does it appear
in the subject index of the 1999 annual conference of the American Educational
Research Association. A tentative explanation could be that school evaluation in
American restructuring mostly appears as assessment of pupil achievement that is
accountability oriented and shaped at state or district level, rather than school
level. I have no explanation, however, as to why a more process oriented type of
school evaluation apparently is not part of site-based management in restructuring
American schools.

2. DEFINITION

2.1. Evaluation

Evaluating means judging the value of an object, and evaluation in the
sense of a particular type of disciplined inquiry emphasizes that this “judging” and
“valuing” is based on some kind of systematic information gathering approach.

In the case where this systematic information gathering is formalized
according to the criteria for social scientific inquiry the term evaluation research
is appropriate. A third major component of evaluation, next to the valuing aspect
and the systematic approach to information gathering, is the applied context:
evaluation results are expected to be used by relevant audiences. Again there
is a prototype situation, often related to policy-evaluation, where evaluation
results are expected to shape, or at least have a certain impact on, policy
decisions.

In the evaluation literature authors vary in their emphasis of each of these
three basic components: valuing, systematic inquiry and use for decision
making. In all types of definitions where goal attainment is centrally placed, the value aspect is prominent (since whether or not program goals are attained provides the basis for judging it as either successful or unsuccessful). Thus, Tyler defines evaluation as "The process of determining to what extent educational objectives are actually being realized" (Tyler, 1950, p. 69, cited by Nevo, 1995, p. 10).

Also Provus' "Discrepancy Evaluation Model" (Provus, 1971) depends heavily on pre-established goals which serve as a basis for judging the success of a program. Scriven's reaction, namely his idea of "Goal Free Evaluation" (Scriven, 1967), also emphasizes the valuing aspect, although he denounces program goals as providing the basic orientation for making judgments. Instead of goals and objectives the demands and needs of clients or relevant audiences of the program that is to be evaluated are seen as the basis for choosing evaluation standards (i.e. the norms used to determine "success" or "failure" of a program).

In making other definitions in the literature both elements of "valuing" and "systematic inquiry" are present, like for example in the definition presented by the "Joint Committee on Standards for Evaluation", led by Daniel Stufflebeam: "evaluation is the systematic investigation of the worth or merit of some object" (Joint Committee, 1981, p. 12, cited by Nevo, 1995, p. 10).

Finally there is a category of authors who seem to altogether leave out the judgmental component from their definitions of evaluation and define evaluation in terms of providing information for decision making.

Stufflebeam's earlier CIPP-model (Stufflebeam, 1983) is an example of this as are authors who speak of "utilization focused evaluation" (Alkin et al., 1979; Patton, 1978). It could be argued that in these approaches the judgmental component is merely left implicit, since valuing is always there whenever information is interpreted as favoring or disfavoring a particular decision alternative.

Apart from these mainstream distinctions in defining (educational) evaluation, there are examples in the literature where still other aspects of the "evaluation endeavor" are centrally placed. Cronbach and his associates, for example, depict the evaluator as an "educator", who enters a dialogue with the professionals in the object situation of the evaluation. Their view is also seen as an example of suppressing the role of the evaluator as a "judge" (Nevo, 1995, p. 10). This view where the component of descriptive information gathering is placed central upon which "illumination" or "education" of the evaluatees and clients is supposed to follow is also predominant in perspectives on evaluation where qualitative description and naturalistic methods are propagated (i.e. Stake, 1975; Guba & Lincoln, 1982).
In what is called "stakeholder-based evaluation" the fact that it is often the case that different parties have an (often divergent) interest in program outcomes, is used in shaping the evaluation. The idea is that giving these various parties more proprietary feeling for the evaluation process and its outcomes will increase the chances of the evaluation results being used (cf. Scheerens, 1990, p. 38).

In advocacy oriented or "judicial" evaluation varying value positions of relevant parties are also used, but more in the final stage of interpreting, and weighing and judging the information that has been gathered (Wolf, 1990, pp. 79–81). During a public presentation of the data a hearing is organized according to the format of the functioning of a court of law. Witnesses are called to provide evidence before or against the case (i.e. the success or failure of a program) and juries decide.

In summary, it seems wise to contain all three elements: systematic inquiry, judgement, and use in decision-making settings in our definition of educational evaluation. Therefore the following definition of educational evaluation is chosen: *Judging the value of educational objects on the basis of systematic information gathering in order to support decision making and learning.*

From the brief overview of views on the evaluation phenomena in the relevant literature it has also become clear that there are some important "contextual conditions" at stake when we deal with educational evaluation. The most important dimension on which these conditions manifest themselves is the variation in positions and interests in the evaluation process and outcomes of relevant parties. This realization gives cause to paying considerable attention to the political and organizational context of school evaluation.

2.2. School Evaluation

In the definition of educational evaluation in the preceding paragraph "educational objects" were referred to. When "schools" are the educational objects to be evaluated – instead of programs in which many schools, teachers or individual students take part – one can speak of "school evaluation".

The fact that schools are the objects which – on the basis of systematic information gathering – are being judged, leaves open the possibility that data on "objects" or "units" within the school are the focus of data collection. However, information on these within-schools units (classrooms, teachers, departments or pupils) will then be aggregated to the school level in order to allow for judgements on the individual school. Such judgements will often require information on other schools, as a basis for comparison.
2.3. Internal and External School Evaluation

There are four main categories of actors in all types of evaluation, including school evaluation:

A. the contractors, funders and initiators of the evaluation;
B. the (professional) staff that carry out the evaluation;
C. the persons in the object-situation which provide data;
D. the clients or users or audiences of the evaluation results.

Mostly categories A and D will partly overlap, in the sense that contractors will almost always be "users" as well, although they may not be the only category of users. For example, a particular department at the Ministry of Education may be contractor and user of a particular program evaluation, although other important parties, such as Members of Parliament and the tax-payers may also be considered as relevant audiences.

If all of these audiences are situated within the organizational unit which is the object of evaluation we speak of internal evaluation (cf. Nevo, 1995, p. 48).

External evaluation occurs when contractors, evaluators and clients are external to the unit that is being evaluated. If functions A, B and C are carried out by the school but the results are mainly used for external clients (D), we could still speak of internal evaluation. Also, when the professional staff is external, but all other categories are internal, the evaluation could still be seen as internal.

2.4. School Self-Evaluation

After the preliminaries in the preceding section, it is now simple to define school self-evaluation, namely as the type of internal school evaluation where the professionals that carry out the program or core-service of the organization (i.e. teachers and head teachers) carry out the evaluation in their own organization (i.e. the school).

This definition would also apply if school teams would make use of external advisors to provide them with council on evaluation methods etc., because the school teams would still take the responsibility for carrying out the evaluation.

The definition of school self-evaluation is analogue to the following definition of "self-report", stated by Newfield (1990): "Self-report refers to the result produced by any measurement technique in which an individual is instructed to serve both as assessor or observer and as the object of the assessment or observation" (Newfield, 1990, p. 146).
3. A FIRST DISTINCTION BASED ON THE EXTERNAL vs. INTERNAL ORIENTATION OF SCHOOL EVALUATION

School self-evaluations may vary to the degree that they are “spin-offs” of external evaluation or entirely internally determined. The following categories can be distinguished, varying from external, to internal.

(a) School self-evaluations that are spin-offs of national or district-level assessment programs, where school results are fed back to individual schools.
(b) School self-evaluations that serve internal and external purposes and are subject to meta-evaluation by inspectorates.
(c) School self-evaluations that are explicitly aimed at providing information to external constituencies as well as aimed at use of the information for school improvement processes.
(d) Self-evaluations that are part of improvement programs that involve a number of schools (evaluations may have the additional purpose of assessing the effects of the school improvement project as a whole).
(e) Tailor-made self-evaluations of individual schools.

West and Hopkins (1998) provide a further qualification of the evaluation orientation with respect to school improvement. They distinguish:

- **Evaluation of school improvement.** In this case the outcomes of improvement efforts or the fidelity of process implementation are the focus. The school evaluation has a summative orientation.
- **Evaluation for school improvement.** In this case evaluation is used during the process of school improvement in order to further shape this process. The orientation is formative rather than summative.
- **Evaluation as school improvement.** In this case the evaluation and improvement processes are one and the same. Perhaps the term “action research” best expresses this orientation to school self-evaluation. I would interpret it as exploiting the reflexive potential of the evaluation processes. For example, the mere exercise of school teams considering the priorities and methods of a search for the strong and weak points of the school’s functioning may lead to improvement in the sense of increased awareness of educational goals and cooperation between staff.

Figure 1 combines the five external/internal orientations with the distinction in evaluation orientation that West and Hopkins make between evaluation of, for and as school improvement.
<table>
<thead>
<tr>
<th>External vs. Internal orientation</th>
<th>Distinction vs. integration of evaluation and improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>School self-evaluation as spin-off of external school evaluation</td>
<td></td>
</tr>
<tr>
<td>School self-evaluation for internal and external purposes, monitored from a central level (i.e. inspectorate)</td>
<td></td>
</tr>
<tr>
<td>School self-evaluation for internal and external purposes</td>
<td>Evaluation of school improvement; one design for several schools</td>
</tr>
<tr>
<td>School self-evaluation as evaluation of school improvement programs involving more than one school</td>
<td></td>
</tr>
<tr>
<td>Tailor-made self-evaluations for each school</td>
<td>Evaluation of school improvement (one school)</td>
</tr>
<tr>
<td></td>
<td>Evaluation for school improvement (formative, one school)</td>
</tr>
<tr>
<td></td>
<td>Evaluation as school improvement (action research, one school)</td>
</tr>
</tbody>
</table>

*Fig. 1. School Self-Evaluation Categories Determined by External vs. Internal Orientations and the Type of Association of School Evaluation and School Improvement.*

**4. A DISTINCTION BASED ON PERCEPTIONS OF SCHOOL QUALITY**

**4.1. What is Quality?**

School self-evaluation activities have the function of monitoring quality. Having said this, one is faced with the task of clarifying what is meant by quality in the sense of “the quality of schools” and “the quality of education”.

First, within the context of school self-evaluation, the interest in quality refers to the school as a whole and not just to the quality of certain aspects or elements like: teaching methods, teachers or school management.
Next, when it is established that school self-evaluation ideally should look at the sum of all aspects and elements of school functioning, one is faced with the need to make certain selective choices and set priorities, if only for practical purposes. But, in order to make these choices, one needs frameworks and/or analytical schemes to capture the "whole" of school functioning. Two of these conceptual frameworks will be used to elucidate choices with respect to quality: (i) a basic model from systems theory; and (ii) perspectives on organizational effectiveness.

4.2. A Basic Model from Systems Theory

An abstract way to picture the functioning of an organization is the construct of the organization as a black box into which inputs flow and by which outputs are "somehow" produced (see Fig. 2).

Even this rather simple model can be used to make the construct of "quality" more concrete. The economic construct of efficiency is defined as the ratio of outputs to inputs, where output (in the case of schools) can be defined in terms of attainment level averaged over pupils, whereas, from the economic perspective, inputs can best be thought of in terms of financial and material resources. Economic efficiency, as a particular operationalization of organizational quality, is focused on the highest possible level of outputs for the lowest cost level of inputs.

In Fig. 2 it is assumed that within the black box processes take place that transform inputs into outputs. When it is attempted to further describe these processes in terms of which process characteristics are most effective in obtaining desired levels of outputs, the model of Fig. 2 becomes more elaborate. In addition, a further distinction of the "input" category is usually made by separating direct inputs into the system and characteristics of the larger context from which these inputs originate. In this way a Context-Input-Process-Output model is obtained. This model is often used as a conceptual framework to summarize the results of school effectiveness research. In Fig. 3 an example of such an ordered summary is shown (cf. Scheerens, 1989).

\[
\text{input} \rightarrow \text{organization as a black box} \rightarrow \text{output}
\]

*Fig. 2. The Organization as a Black Box.*
The notion of quality inherent in integrated school effectiveness models like the one in Fig. 3 is that:

(a) outputs are the basic criteria to judge educational quality;
(b) in order to be able to properly evaluate output, achievement or attainment measures should be adjusted for prior achievement and other pupil intake characteristics; in this way the value added by schooling can be assessed;
(c) in selecting variables and indicators to assess processes and context one should look for those factors that have been shown to be correlated with relatively high "added-value" factors.

It should be noted that school effectiveness models do not prescribe the types of outputs that should be used to assess quality. In principle all types of outputs, cognitive or non-cognitive could be inserted in the right-hand box of Fig. 3. In the actual practice of school effectiveness research cognitive outcomes, mostly in terms of achievement in core-subjects like reading, arithmetic, and language, have predominated. The process factors shown in the middle section of Fig. 3, might well be somewhat different if non-cognitive outcomes or less subject-matter tied cognitive outcomes would have been used in the actual research studies.

It should also be noted that there is still quite a lot of uncertainty about the selection of process factors such as indicated in the figure. The available knowledge-base is far removed from a situation where it would be possible to make precise predictions on the likely added value of schooling, given the state of certain processes.

To the degree that educational effectiveness models provide an acceptable operational definition of quality, they can also be used as a guideline in the design of instruments for school self-evaluation; points a (focus on outcomes); b (proper adjustment of outcomes); and c (measure process characteristics associated with high added value) mentioned in the above can be read as as many guidelines to make choices with respect to instrumentation.

However, a broader perspective on quality can be considered. Such a broader perspective can be obtained from multiple orientations towards organizational effectiveness that will be discussed in the next section.

4.3. Multiple Criteria to Assess Organizational Effectiveness

In organizational theory models like the school effectiveness model are seen as belonging to just one of several effectiveness perspectives. The effectiveness
School Self-Evaluation

Context
- achievement stimulants from higher administrative levels
- development of educational consumerism
- 'covariables,' such as school size, student-body composition, school category, urban/rural

Inputs
- teacher experience
- per pupil expenditure
- parent support

Process
school level
- degree of achievement-oriented policy
- educational leadership
- consensus, cooperative planning of teachers
- quality of school curricula in terms of content-covered and formal structure
- orderly atmosphere
- evaluative potential

classroom level
- time on task (including homework)
- structured teaching
- opportunity to learn
- high expectations of pupils' progress
- degree of evaluation and monitoring of pupils' progress
- reinforcement

Outputs
student achievement adjusted for:
- previous achievement
- intelligence
- SES

Fig. 3. A Summary of the Findings from School Effectiveness Research, from Scheerens, 1989.
perspective in which the school effectiveness model fits is referred to as the *rational goal model*, where productivity and efficiency are the central criteria. Alternative models are: the *open systems model*, with growth and resource acquisition as effectiveness criteria; the *human relations model* with human resource development as a central criterion and the *internal process model* in which stability and control are the main issues. Quinn and Rohrbaugh (1983) depict these four models as determined by two dimensions; one that has flexibility and control as its extremes and one that represents an internal vs. an external orientation (see Fig. 4).

From this framework additional process indicators of school functioning may be generated.

As far as the rational goal model is concerned, it should be noted that this model does not specify which educational objectives are relevant. Next to knowledge and skills in basic school subjects, other educational aims may be recognized. Two important additional categories of educational objectives are *social, emotional and moral development* on the one hand and the development of *general cognitive skills* on the other. For our purposes these categories of educational aims (next to the basic cognitive skills that have been the focus in empirical school effectiveness research) are relevant to the degree that they may require somewhat different teaching approaches and different school organizational arrangements than the process variables that have been shown to matter in the traditional school effectiveness models (Scheerens, 1994).

According to Goodlad and Anderson (1987) *multiage* and *interage grouping* have the advantage of fostering social and emotional development apart from being effective in realising traditional educational goals. The disadvantages of a competitive achievement-oriented atmosphere are supposed to be modified by these organizational arrangements, while the motivational disadvantages of both promoting and non-promoting as in graded system are prevented. *Non-gradedness* and *team-teaching* are seen as measures to realise differentiated adaptive teaching and an integrated, continuous learning route. Such approaches are thought to contribute to the degree that students are comfortable and happy in the school.

Educational psychologists increasingly emphasize the importance of self-regulated learning and meta-cognition. “Subject-free” cognitive skills can be acquired in programmes in which learning how knowledge is acquired (“learning to learn”) is taught.

The *human relations model* is strongly concerned with the work satisfaction of teachers. Louis and Smith (1990) identified seven “quality of work life indicators”: 
- respect from relevant adults, such as the administrators in the school and district, parents, and the community at large;
- participation in decision making that augments the teachers' sense of influence or control over their work setting;
- frequent and stimulating professional interaction among peers (e.g. collaborative work/collegial relationships) within the school;
- structures and procedures that contribute to a high sense of efficacy (e.g. mechanisms permitting teachers to obtain frequent and accurate feedback about
their performance and the specific effects of their performance on student learning;

• opportunity to make full use of existing skills and knowledge, and to acquire new skills and knowledge (self-development); the opportunity to experiment;

• adequate resources to carry out the job; a pleasant, orderly physical working environment;

• a sense of congruence between personal goals and the school’s goals (low alienation).

Other factors that may contribute to teachers’ satisfaction are task differentiation and possibilities to make promotion (though these are usually limited) and financial incentives, though this approach, according to some authors might prove counter-productive – McLaughlin & Mei-ling Yu (1988).

The open system model emphasizes the responsiveness of schools with respect to environmental requirements. This means on the one hand that school organizations can create effective buffers against external threats, and on the other hand that schools can manipulate their environments to the degree that their own functioning is not only safeguarded but also improved. In some countries (The Netherlands for instance) external regulations for schools are relaxed and school autonomy is enhanced. This state of affairs offers new possibilities, but also confronts the school with new requirements, for instance to conduct their own financial policy.

Demographic developments (less pupils) may force schools to be active in stimulating student enrolment and “school marketing.” Developments in educational technology, initiatives for educational innovations from higher administrative levels as well as accountability requirements can be seen as additional external forces that challenge the school’s readiness to change.

In a Dutch study, Gooren (1989) found evidence for a dichotomy of schools that could either cope or not cope with these new external requirements. The schools that could cope more frequently had strong leadership or a collegial structure in contrast to non-coping schools which answered the image of the loosely-coupled, segmented school organization.

Capacities of schools to deal with an increasingly demanding and dynamic environment are described in terms like “the policy-making potential of school” and “the self-renewing capacity of schools”. School organizational characteristics that are thought to contribute to these capacities are:

• leadership (also in sense of entrepreneurship);

• collegiality;
School Self-Evaluation

- capacity for self-evaluation and learning (see for instance Morgan’s image of the learning organization – Morgan, 1986, Ch. 4);
- overt school marketing activities;
- strong parental involvement;
- boundary-spanning positions;
- support of external change agents.

Proxy-indicators concerning the success of responsiveness are enrolment figures and characteristics of buildings and equipment.

Whereas the human relations model is concerned with social and cultural aspects of “what keeps organizations together,” the internal process model reflects a preoccupation with formalization and structure. From this perspective the following factors are of interest:

- explicit planning documents (such as school curricula, school development plans);
- clear rules regarding discipline;
- formalization of positions;
- continuity in leadership and staffing;
- integrated curricula (coordination over grades).

Proxy-indicators for the stability of school organizations are attendance rates, the number of teaching periods not given, and figures about the continuity in staffing.

4.4. Quality Indicators

The ideas for additional process indicators that come from this more comprehensive treatment of organizational effectiveness are summarized in Fig. 5. (Process indicators induced from the narrower model of school effectiveness research are also included.)

5. A TAXONOMY OF BASIC TYPES OF SCHOOL SELF-EVALUATION APPROACHES

In this section, basic types of school self-evaluation will be discussed. By “basic” we mean that these approaches, which are mainly distinguished on the basis of evaluation methodology, have a certain tradition and are rooted in
Human relations model

* Quality of work life indicators
  * respect
  * participation in decision making
  * professional interaction
  * performance feedback
  * opportunity to use skills
  * resources
  * congruence personal/organizational goals

Open system model

* entrepreneurship
* collegiality
* capacity for self-evaluation and learning
* overt school marketing activities
* parental involvement
* boundary-spanning positions
* external change agents
* student enrolment figures
* resources (buildings, equipment)

Internal process model

* planning documents
* disciplinary rules
* management information systems
* formalization of positions
* continuity in staffing and leadership
* integrated curricula
* attendance rates
* lessons “not given”

Rational goal model

* (school effectiveness research)
  * educational leadership
  * success-oriented ethos
  * monitoring of student’s progress
  * time on task
  * content-covered (opportunity to learn)

* (broader set of educational goals)
  * non-gradedness
  * team teaching
  * individualization, differentiation
  * continuous learning route
  * time spent on social, emotional, creative and moral development
  * “learning to learn” activities
  * diagnostic testing

Fig. 5. Additional Factors for Process Indicators Generated from the Quinn and Rohrbaugh Framework.
specific social-scientific disciplines. In later chapters, additional, more recent approaches will be discussed as well.

Stating that these approaches are "basic" and perhaps more "classic" than some of the methods to be discussed further on does not mean that their application should be considered "less innovative". Even for these approaches there exists no widespread practice and application, particularly from an international perspective. Indeed, pupil monitoring systems and school management information systems often require very sophisticated instruments and tools and are therefore potentially innovative from a more technical point of view. School-based review approaches are likely to be innovative from a different perspective, namely in the challenge they provide with respect to the social functioning of school teams, the discussion of norms and values and what is sometimes referred to as "organizational learning" by teachers as "reflective practitioners" (Argyris & Schôn, 1974).

5.1. Basic Types of School Self-Evaluation

Currently, several approaches to school self-evaluation are being used. Each has a specific disciplinary background and a specific context in which the approach was originally employed, as is shown in Table 1. Each of these approaches will be sketched briefly and strong and weak points will be discussed.

School-Based Review
School-based review depends heavily on opinions of school personnel on discrepancies between the actual and an ideal state of affairs in schools. In this way a broad perspective, in which all the main aspects of school functioning can be scrutinized, is possible. Usually, respondents are also asked to indicate

<table>
<thead>
<tr>
<th>Approach</th>
<th>Disciplinary Background</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>school-based review</td>
<td>social psychology, education</td>
<td>schools</td>
</tr>
<tr>
<td>management information systems</td>
<td>business administration, operations research</td>
<td>private industry</td>
</tr>
<tr>
<td>organizational diagnosis</td>
<td>management consultancy</td>
<td>public-sector organizations</td>
</tr>
<tr>
<td>pupil monitoring systems</td>
<td>educational measurement</td>
<td>(remedial) teaching</td>
</tr>
</tbody>
</table>
whether a certain discrepancy should be actively resolved. This approach to school self-evaluation seeks to gear improvement-oriented action to appraisal. The context of application is usually school improvement, which means that a school-based review is carried out when there is a prevailing commitment to educational innovation.

Advantages of this approach are: a broad scope, a user-friendly technology, an explicit linkage between evaluation and action, and a high degree of participation (all school personnel take part in the review). A definite weakness of school-based review is its dependence on subjective opinions and its (usual) neglect of “hard” factual data on school functioning, most notably output data.

Examples of procedures for school-based review are the GRID and GILS-systems (see Hopkins, 1987) and the SAS-system (Voogt, 1989).

**School Management Information Systems**

School management information systems have been inspired by similar systems in private industry. Generally they consist of a careful modelling of information streams and information needs within a company, deciding which data should be available for the purpose on a more or less permanent basis, followed by design and implementation of a computer configuration and software. Bluhm and Visscher (1990) describe a management information system as an information system based on one or several computers, consisting of a data-bank and one or several software applications, which enable computer-based data storage, data analysis and data distribution.

A question that could be answered by means of such a school management information system would be: “to which degree has absenteeism decreased after the implementation of specific measures to fight absenteeism?”

Management information systems have a great potential for supplying important information on a routine basis. At present practical barriers; one needs to have sufficient and adequate computer hardware and even when professionally developed software packages become available, quite a few specific maintenance functions must be carried out, while new routines and perhaps even functions to guarantee adequate data-entry should be developed.

**Educational Indicators**

Although educational indicator systems are usually employed at the macro level (the level of national educational systems), for instance to describe the “state of education” of a country on a yearly basis, some authors have suggested
applications at the school level (Teauber, 1987; Oakes, 1987; Scheerens, 1990). When applied at the school level, educational indicator systems typically will include “process” or “throughput” information, next to input, school-context and output data.

Results of school effectiveness research studies are usually employed to select process indicators. The general idea of indicators is to provide an at-a-glance profile of certain important characteristics of an educational system.

This means that there is no aspiration to “dig deep”, while employing easily measured characteristics and so-called proxy measures. This feature is at the same time a definite limitation of the approach. The “indicator” approach at school level differs from school management information by a more substantive educational, rather than an administrative orientation. The two approaches are quite close to one another, however, as there may be overlapping areas, like, for example, the efficient use of time at school and classroom level. As compared to school diagnosis the indicator approach has a quantitative rather than a qualitative orientation and a stronger preoccupation with collecting objective data rather than subjective descriptions and opinions.

In Fig. 1 the various categories on the internal/external dimension are combined with Hopkins and West’s (1998) distinction of three types of associations of evaluation and improvement. It should be noted that Hopkins and West consider only evaluation and improvement activities that are unique for one school. In Fig. 1 evaluation of school improvement programs using one design for several schools is added as a relevant category.

Organizational Diagnosis

As educational institutes (schools and universities) are made to function more autonomously, it is quite likely that they will become more like private companies in their managerial and organizational characteristics. An example of this would be a stronger emphasis on strategic planning and on scanning the external environment of the school. It is therefore not surprising that approaches used in management consultancy are introduced in schools. Although these approaches, generally labelled as “organizational diagnosis” or “management audit”, usually depend on an external organizational consultant – they are also available for school self-diagnosis. In contrast to school-based review these approaches tend to be exclusively based on information provided by the management of the organization. So, when they are used without an external consultant they would appear to be somewhat like “management introspection.” A strong point of this approach is that it is likely to pay attention to issues that were kept largely unnoticed by the educational province, such as external contacts, anticipation of developments
in the relevant environment, and flexibility in offering new types of services. The most important disadvantage remains, however, that this approach is not so easy to transform to a school-based application, without an external consultant.

**Pupil Monitoring Systems**

The focus of attention in this article is self-evaluation at the school level. Pupil monitoring systems operate at the micro level (class level) of educational systems. In the ensuing sections of this article it will be shown how this class of techniques can also be used for self-evaluation at the school level.

Basically, pupil monitoring systems are sets of educational achievement tests that are used for purposes of formative didactic evaluation. An important function is to identify those pupils who fall behind and where they experience difficulties.

Pupil monitoring systems have one asset which, in our opinion, is essential for all efforts to make school functioning more effective: the centrality of output data at the level of the individual pupils measured by means of achievement tests. If approaches to school self-evaluation neglect these type of data there is a risk that the information basis they supply for educational or administrative decision making is faulty (see the earlier reference to the phenomenon of goal displacement).

5.2. Basic Types of School Self-Evaluation and Perspectives on Educational Quality

The four types of perspectives on organizational quality, distinguished by Quinn & Rohrbaugh (1983) and presented in Section 1.2, can be related to the basic types of school evaluation as distinguished in the preceding paragraph (1.3.1).

The association of organizational quality perspective and self-evaluation approach is based on the similarity in disciplinary orientation and correspondence in the criteria that are likely to be central in each of the evaluation approaches, see Fig. 6.

6. METHODS AND INSTRUMENTS

6.1. A Taxonomy Distinguishing Methods, Actors and Objects

When school evaluation at large – not exclusively school self-evaluation – is considered and when methods are distinguished on the basis of actors and
Fig. 6. Association of Organizational Effectiveness Perspective and Basic School Self-Evaluation Approaches.

objects of the evaluation a more extensive set of approaches can be distinguished:

**Evaluation methods, when pupils are the object**
- informal procedures of evaluating learning tasks, marking [teachers];
- curriculum-tied progress tests for different subjects (i.e. unstandardized tests) [teachers];
• semi-formal presentations of completed learning tasks such as portfolios [teachers];
• authentic assessment, i.e. when pupils’ progress is evaluated in natural circumstances [teachers, schools];
• pupil monitoring systems of standardized tests and assignments [schools];
• certifications (not necessarily with diploma) [central government];
• assessment tests initiated by [local, regional or national authorities].

Evaluation methods when teachers are the object
• formal methods of teacher appraisal [school boards, school leaders, inspectors].
• informal methods of teacher appraisal [school boards, school leaders]
• evaluating teachers by means of observation of the quality of instruction [senior school management]
• ratings of instructional quality by students [students]

Evaluation methods when the school (or department within a school) is the object
• school diagnosis in the form of so-called “GRIDS” depending on opinions and self-appraisal of school staff [school leader, department];
• school management information systems, e.g. a computerized registration of absenteeism [school management and other administrative levels];
• integrated school self-evaluation systems in which assessment of school processes is combined with assessment of pupils’ achievement [school management, head of department];
• so-called “visitation committees”, whereby peers (e.g. colleagues from other schools) screen and evaluate a school [unions of schools];
• accreditation, whereby an external private company screens aspects of school functioning using a formal set of standards [private agency];
• inspection, qualitative or semi-qualitative assessment by inspectors of school [Inspectorate];
• school level indicators or key data (school monitoring) [school management and other administrative levels];
• assessment and market research of the school in its relevant environments, e.g. with respect to expectations on future enrolments [external research institute];
• external school review by [private consultance institutes]
Evaluation methods when the system of schools is the object:
• national assessment [national government];
• program evaluation [national government];
• inspection [national government];
• educational indicator projects [national government]

6.2. A Closer Look at Specific Types of Instruments

The range of data-collection methods and types of instruments for school self-evaluation comprises virtually all that is available in educational research, and educational evaluation and assessment.

Since it is an impossible and – given the availability of Handbooks on each of these methodologies – also an irrelevant task to try and give an exhaustive overview – only a few specific types of instruments and data collection methods will be briefly discussed. Vertically equated achievement tests used in pupil monitoring systems, student ratings of teachers’ instruction and photo-evaluations will be considered as examples of the wide range of available methods.

6.2.1. Vertically Equated Achievement Tests

Ideally, pupil monitoring systems make use of achievement tests that confirm to the assumptions of item response models. One of the advantages of these psychometric models is that the results on different tests that measure the same domain of items with varying degrees of difficulty can be compared. Such tests are ideal to measure progress as pupils advance through grade levels. Another advantage is that such tests are criterion-oriented, in the sense that it is clear to which specific content domain they refer. This means that it is possible to assess at various points in time, in which sub-domains pupils are strong or weak.

In the pupil monitoring system that is used in the Netherlands, tests are administered three times a year, at each grade level (Scheerens, Hendriks & Bosker, 1998). For school self-evaluation purposes test results can be aggregated to class and school-level and possibly – when links can be made to school management information systems – adjusted for the socio-economic background of pupils.

Another use made in the Dutch Integrated School Self-evaluation Project (Scheerens, Hendriks & Bosker, 1998) of the set of achievement tests used in the pupil monitoring system consists of teacher ratings of the degree to which pupils were expected to be capable of scoring the items correctly. These ratings were used as a measure of opportunity to learn, i.e. the degree of overlap between subject matter elements taught and tested.
In the Netherlands the pupil monitoring system (developed by CITO) is used, on a voluntary basis, by 70% of the primary schools. It is still mainly used for diagnosis of the progress of individual students, while the use for school self-evaluation is only just beginning.

6.2.2. Pupils Rating Teachers, and Teachers Rating Principals

As an alternative to self-reports that are frequently used to measure "process" characteristics of schooling like educational leadership and teaching strategies, ratings by other "actors" are increasingly considered in the context of school effectiveness research and school self-evaluation (cf. Hill, Rowe & Jones, 1995; Scheerens, Hendriks & Bosker, 1998).

Evidence on the reliability and validity of a short-scale on which primary school pupils evaluated their teachers is provided by Kuyper and Swint (1996). Van Os (1999) showed that ratings of the instructional behavior of teachers in higher education had an acceptable level of reliability and validity.

Figure 7 provides an example of the way the results of a rating procedure (teachers rating head teachers) in Dutch secondary schools are presented to schools.

An interesting finding in Van Os' study was that, contrary to common suspicion, students' judgements about their teacher were not significantly

![Example of feedback school report](image)

From the school overview it appears that teachers from the example school in comparison with teachers from other schools are above-average positive about the educational leadership at their school. The workload compared with other schools is below average and all teachers agree with this (see width of white bar). The scores for staff development are somewhat lower than those from other schools. From the school overview we can see that the average score on cooperation at the example school is very low and that the teachers have diverse opinions about this subject (see width of white bar).

**Fig. 7.** Example of the Feedback About the School as a Whole (From Scheerens, Hendriks & Bosker, 1998).
influenced by their achievement levels. Apart from methodological advantages as compared to self-reports, having different actors describe and rate each others behavior also has considerable action potential, particularly when ratings by others are compared to self-reports.

6.2.3. Photo Evaluations as an Example of Qualitative Methods
Qualitative methods are distinguished by a non-random selection of (usually a small number of) units of study, open, unstructured procedures for collecting data, use of narrative and “thick” description and, sometimes, “participation” of the persons in the object-situation in all aspects of the study process.

MacBeath (1999) discusses case studies, self-monitoring (“Evaluation of how well my lesson went”), diaries and logs, spot checks (e.g. detailed descriptions of what the students are doing at a given short period of time), photo and video registrations of significant places, events and people, shadowing (as when a student is followed throughout a school day), open interviews, focus groups, using an “urgency and importance matrix”, development analysis interviews, whereby different stakeholders state their point of view on a particular issue and critical incident techniques (detailed unpacking of a significant event).

The following citation from the Newsletter of the “European Pilot Project on Quality Evaluation in School Education” of November 1998 illustrates photo evaluation:

Camera in hand, the pupils went around the various parts of the school and took photos of one another. They then gave their judgements, impressions, feelings about what they experienced in those particular spots. What we were interested in was not so much their views on the aesthetic or the functional aspects, but how they felt in those places: happy, sad, depressed, at ease, ill at ease, etc. An interesting overall picture emerged and the results were put on public display as a photo exhibition. The pupils showed us places we were unaware of, secret spaces. Their simultaneous need for protection and independence emerges clearly. They want congenial, not threatening surroundings; a place which also allow them to be alone, to separate off, to think, without hustle and bustle, without noise. The information greatly improved our understanding of them. During the second phase, some classes worked on the design of places, the garden, the internal spaces, the classrooms, using the design workshop organised in conjunction with the School of Architecture at the University of Florence.

Rosaria Bortolone, Headmistress of S. M. S. Piero della Francesca-Pirandello, 
Florence, Italy.

Qualitative approaches to school self-evaluation indicate that there is only a gradual difference between disciplined inquiry and the work of teachers as reflective practitioners, or rather that both approaches gradually blend into each other.
Strengths of qualitative approaches lie in their closeness to every-day practice and their open, participative nature. Weaknesses are: representativeness, objectivity and efficiency. This last aspect efficiency refers to the labor-intensity of detailed, “thick” description.

7. IMPLEMENTATION ISSUES; APPLICABILITY IN DEVELOPING COUNTRIES

In this final section feasibility of implementation of school self-evaluation approaches will be considered. Again the evidence is based on experiences in Europe. In particular the results of three research projects funded by the European Commission will be used, these are the EEDS-project (Evaluation of Educational Establishments – Van Amelsvoort et al., 1998); the INAP project (Innovative Approaches to School Self-Evaluation – Tiana et al., 1999) and the EVA-project (Quality Evaluation in School Education – e.g. Hingel & Jakobson, 1998). All three projects provide extensive information on case-studies of school self-evaluation activities in European countries.

7.1. Reconsideration of the Internal/External Dimension

The EEDS and INAP projects found that in practically all cases that were studied in five countries (Scotland, England & Wales, Spain, Italy and the Netherlands) there was a strongly external impetus to the school evaluation projects that were studied. The projects that were studied were usually hybrid forms in which external and internal elements were both present. In all cases networks of schools collaborated in school (self-)evaluation activities. Mostly initiatives came from above school units, municipalities, local education authorities or regional support agencies. In all cases schools obtained external support and mostly used externally developed instruments. In a minority of cases schools adapted externally developed instruments or developed their own instruments with the help of external experts.

The evidence from the EVA-project illustrates genuine school-based initiatives more frequently, although external support is usually present in these cases as well.³

The reality of school self-evaluation, particularly in countries where this practice is a very recent phenomenon, is “external evaluation with an increasing degree of school participation” rather than genuine school self-evaluation. So far, the most common initiation and implementation strategy in Europe seems to be “spin-off” from externally initiated types of school evaluation.
Nevertheless, there are other examples that are more genuinely school-based as well. The example of Dutch primary schools that buy their own pupil monitoring system, and which was referred to earlier, is a case in point. There are also some very positive experiences where schools work with external experts on setting priorities and standards for school self-evaluation (MacBeath, 1999; Scheerens, 1999). These latter examples are tending towards what West and Hopkins describe as evaluation as school improvement.

The relevance of these experiences for developing countries is twofold:

Firstly, school self-evaluation can be initiated very well by exploiting the spin-off of external evaluations, like national monitoring systems or evaluations of development projects. Pre-requisites for such practice are that information is available at lower levels of aggregation (schools, classrooms) and that specific measures are taken to feed this information back to schools in a comprehensible way.

Secondly, the introduction of basic and simple forms of school self-evaluation in schools in developing countries can be used as a feasible and practical way to bring about a process of self-reflection and school improvement. This latter practice, however, would require a local cadre of support staff, e.g. an inspectorate.

7.2. External Support

In all cases described in the EU-studies there was some kind of external support for the schools that participated in school self-evaluation projects. The type of required support, as a matter of course, depends on the type of school self-evaluation that is chosen. There are two main areas of support: technical support and management support in creating and maintaining the organizational conditions required for an effective use of self-evaluation. In cases where self-evaluation is largely a spin-off of external evaluations, involving many schools, data will be processed and analyzed externally. Special efforts will need to be made to feed back data to individual schools in an accessible and comprehensible way. In these situations schools would also require some guidance in the interpretation of results, application of standards and benchmarks.

When the choice and development of evaluation methods is more of a bottom-up process, schools would require some technical guidance in providing a range of possible approaches, methods and instruments and in the technology of instrument development. As stated before, such collaborative activities, to some extent, are school improvement activities in their own right as they urge school teams to collaboratively reflect on major goals and methods of schooling.
Management support is needed to create and maintain organizational conditions necessary to conduct school self-evaluations. In fact the implementation of school self-evaluation is to be seen as an innovatory process, to which all principles of good practice apply. One of these principles is the essential role of the principal. Other aspects are seeking the involvement of all staff and external constituencies. A basic organizational requirement for good practice of school self-evaluation is the institutionalization of some kind of forum where staff can meet to plan evaluation activities and discuss results.

Apart from technical and managerial support, in many situations, schools would also require more substantive educational support in interpreting results and designing remediation and corrective actions to improve the school's functioning in weak areas. There is definitely the danger of creating an overload of evaluative information that is not fully exploited for its action potential. To put it differently, self-evaluation should not end in diagnosis but be actively used for "therapy".

From my (limited) experience with education projects in developing countries it would seem that technical support in analyzing and feeding back data is not the most difficult requirement to meet. The individual support to schools in interpretation of data, participation in the development of instruments and procedures and information use, would be a much more difficult condition to fulfill.

7.3. Cost Aspects

The need for external support and guidance is the more expensive to the degree that each and every school would develop its own "tailor-made" approach to school self-evaluation.

Economies of scale, in working with networks of schools and projects involving many schools, are to be considered, when resources are scarce. School self-evaluation on the basis of data feedback from existing national assessment or monitoring projects exploits this principle even further.

Local support staff to guide schools in school self-evaluation seems to be an unrealistic pre-condition for many developing countries. There would be a lot of potential in small-scale pilot projects, however, where the use of school self-evaluation could be implemented and studied in the specific local context. Among other applications, such experiences could be used in the design of training courses as part of regular training of teachers and head teachers.

Experiments with in-service teacher training activities in school self-evaluation could also be seen as long-term investments in the building of local
capacity in the directly practical and foundational skills that are at stake in creating schools that can handle autonomy and self-improvement.

7.4. The Micro-Politics of Evaluation

Since evaluations – even school self-evaluations – ultimately lead to judgements and “valuing” – some categories of actors, particularly teachers, are likely to feel threatened. Traditionally schools have functioned according to the principles of the “professional bureaucracy” (Mintzberg, 1979), where enculturation and training in the profession is the key control mechanism and autonomous professionals are described as opposing rational techniques of planning and monitoring.

School evaluation activities have the potential of external control in areas which were traditionally safeguarded under the umbrella of the professional autonomy of teachers. The subsequent greater transparency of the primary process of schooling to external parties, e.g. the principal and the school board, has implications for the balance of power within schools. In the early literature on program evaluation clashes between evaluation experts and practitioners have been documented as the confrontation of “two worlds” (Caplan, 1982); and such tensions cannot be ruled out even when evaluation is internal and improvement-oriented. Several authors have therefore emphasized the creation of non-threatening conditions for school evaluation (Nevo, 1995; MacBeath, 1999). The role of the external expert should become more like an advisor and a “critical friend” to the school.

School evaluation can be perceived in a context of accountability and a context of improvement. Theoretically one would expect that evaluation apprehension would be stronger in an accountability as compared to an improvement context. In actual practice, at least in Europe, school self-evaluation often arises as a consequence, spin-off or counter-balance to accountability-oriented assessments. Reconciliation and integration of accountability and improvement orientations is the more likely when the external control element, most notably the taking of sanctions, is less severe. In Europe there are examples where external accountability-oriented assessments, like the production of league-tables, actually function as the main incentive for schools to embark upon school self-evaluation which considers a broader spectrum of aspects of school functioning.

But even when there is no accountability at stake, and school self-evaluations are designed bottom-up, the issue of teachers feeling threatened arises. It is therefore important that school self-evaluation is clearly and explicitly introduced to all stakeholders and participants and that initial activities are
experienced as intrinsically, professionally rewarding. Ultimately the relevance and use of data and application of standards for all school staff should function as the main incentive to sustained school self-evaluation.

The micro-politics of school evaluation are likely to differ according to the structure and educational culture of a country. Therefore, no generally applicable guidelines can be given for applications in developing countries other than the strong recommendation not to overlook the political aspects and all the repercussions they may have for issues of reliable data-collection, anonymity of results, facilitation of coupling databases and good professional cooperation between teachers, principals and support staff.

CONCLUSION

In this paper, school self-evaluation has been defined as a type of school evaluation where the school takes responsibility for its own evaluation. From an extensive overview of categorizations it appears that there are many forms where the school taking responsibility does not preclude the shaping of evaluation methods by external parties. Case-studies from Europe indicate that in many instances school self-evaluation occurs as a spin-off, consequence or counterbalance to external evaluation. Many of these case-studies show an orientation to accountability and self-improvement rather than an exclusive preoccupation with one of each. When it comes to applying school self-evaluation in developing countries the European experience of hybrid forms of external and internal school evaluation is seen as an advantage rather than a handicap. Similarly, from a methodological perspective, integration and combination of different "pure" types of school self-evaluation appears to be the rule rather than the exception.

Given the costs, the required expertise and the fact that in many developing countries system-level assessment and monitoring are already implemented or in a stage of development, school self-evaluation could get off the ground in the wake of these large-scale programs.

Bottom-up developments, where schools design their own self-evaluation, should also get a chance, however. For these, small-scale pilot projects could be set up to explore the possibilities that planning school self-evaluation offers as a form of reflection and school improvement in its own right. Results of such pilots could have an important function in the shaping of initial and in-service teacher training programs.

A final observation – also for the application in developing countries – was that the micro-politics of evaluation should be an important focus of consideration in the way school self-evaluation is introduced and designed. Tackling
School Self-Evaluation

this potential problem area can well avoid a lot of loss of energy in dealing with resistances, distortions and even corruption of evaluation.

School self-evaluation contains the possibility to bridge the distance between evaluation and school improvement, particularly when it is tackled as a joint learning experience from internal and external actors, like administrators, school leaders, teachers and external researchers. It is therefore to be seen as an important lever to educational change and improvement with considerable potential for developing countries.

NOTES

1. Actors are indicated in brackets.
2. In fact, schools have to buy the system and services.
3. The outcomes reflect, to some extent, the focus, or sampling bias of these studies, where EEDS and INAP self-evaluation projects, whereas EVA sampled individual schools in each EU country.

REFERENCES


Os, W. van (1999). *Bruikbaarheid en Effectiviteit van Studentenoordelen over het Onderwijs [Usefulness of Student Ratings]*. Amsterdam: Free University of Amsterdam.


Scheerens, J. (1999b). The STRESA and ELANS project in Italy (manuscript in preparation).


SCHOOL-BASED EVALUATION
IN NORWAY:
WHY IS IT SO DIFFICULT
TO CONVINCE TEACHERS OF
ITS USEFULNESS?

Lars Monsen

INTRODUCTION

In this chapter I will reflect upon the Norwegian experience with school-based evaluation and try to answer the question: Why is it so difficult to persuade teachers to do school-based evaluation? Part of my answer will be found in the different perspectives on evaluation seen from governmental and municipal level on the one side and ordinary teachers on the other. In many schools we have studied, we use a common pattern: Too much pressure to do school-based evaluation creates suspicion of a hidden agenda from above leading to lacking involvement in what should be an important part of school development.

SCHOOL-BASED EVALUATION

School-based evaluation in Norway is only about ten years old. The evaluation of schools is of course older than this term suggests, but it depends on what you mean by ‘evaluation of schools.’ Actually evaluation of schools in Norway...
began more than sixty years ago when educational researchers asked children what they remembered of the different subjects. And in the 70s a government committee proposed additional formal investigations into the output of schools (KUD, 1974). But in practice little has been done to systematically evaluate how schools achieved their goals. In the late 80s this question came to occupy the center stage in the debate on policy.

The OECD made an evaluation of educational policy of Norway. It was concluded that Norway has one of the most decentralized school systems in Europe, but little was known of the effects of this decentralization and other effects of official policy (OECD, 1989). The report led to heated debate politically and among people in the “system,” particularly university academics and educational researchers were engaged debaters. A year before this, a very influential “white paper” had been published about reforms in the university and college sector (KUF, 1988). The committee behind the report was headed by a former university professor in sociology, the then leader of the labor union’s research institute, Gudmund Hernes, who later became minister for the Department of Church, Education and Research. In this report it was stated that the school system in Norway hadn’t achieved its goals in the preparation of the next generation for the knowledge society. They used some research evidence to support their claim, but most of educational researchers taking part in the debate vehemently opposed the committee’s claims. It can therefore be said that the course of educational policy in Norway was made rather unpredictable in the 90s.

On the political front there were many attacks on “the social democratic hegemony” of educational policies from the 50s, this despite the fact that Norway had had some liberal and liberal-conservative governments in the 60s, 70s and the 80s. These attacks were fueled by “the new right” who drew on arguments from England and the U.S.; the claim was that the quality of education in Norway wasn’t good enough. They demanded more and better evaluation of pupils, teachers and the schools as organizations for learning. In this political climate the association for Norwegian private employers, NHO, took an active part in the public debate organizing conferences and seminars and expressing concern for poor standards in the educational system. In the 90s NHO presented several reports with data and arguments in support of a “new evaluation policy” (NHO, 1991). At the government level, the Department for Church, Education and Research, tried to find a better grounding for their policies in order to face the criticism made by the OECD. The report the Ministry made for the OECD, was translated in Norwegian and thereafter a new task force was appointed with the mandate to come up with new ideas for “a national evaluation system.” This group, named EMIL, invited many of the leading evaluation theorists and
researchers in Europe and U.S. to come to Norway to participate in a conference. The conference reports (Granheim & Lundgen, 1990) and proposals to the government were published. In my opinion this group was instrumental in defining a new course in very murky water. In their final report (Granheim & Lundgren, 1990), they recommended that in a national system for evaluation of schools the self-evaluation in schools should be an important part and basis for other kinds of school evaluation. In the meantime other task forces in the government had proposed, and the parliament had voted on, a new system of management by objectives both in education and in other public organizations. In the “white paper” proposing the new system in education, it was stressed that evaluation of goal achievement should be an important part of the educational system.

I have given this background for the introduction of the important new term of the 1990s, school-based evaluation, because as can be seen in this short historical revue, at the start of this decade, it was far from clear what the term “a national evaluation system” really meant. The new ideology of steering by objectives could have led to a new governmental body with responsibility for measuring goal achievement. Instead the government chose to follow up advice from the EMIL group and make a manual for self-evaluation in schools. In this period (1989), my colleague Trond Alvik and I started a project in cooperation with the regional educational office and with support from The Norwegian Council for Research in Science and Humanities, we carried out action research in self-evaluation in schools. In 1991 Alvik published his first book on school-based evaluation. The book and his definition of school-based evaluation provided a more secure position for both academically and policy debates. The same year, Tom Tiller (initiator of another important project in school-based evaluation in upper secondary schools from 1989) and I published a book together about “the efficient school movement” where we tried to point out some of the problems this way of thinking could have for teacher professionalism in Norway (Monsen & Tiller, 1991), and we suggested alternative ways of thinking about school-based evaluation. We all took part in in devising the new manual for school-based evaluation (adopting Alvik’s term) and it was plain that the formulation of this manual involved a difficult political balancing. When it was published in 1993 after more than two years’ work, it determined the policy for years to come (KUF, 1993). In the foreword, the minister (Gudmund Hernes) said that this was a handbook for school-based evaluation and would not prescribe methods for how to do school-based evaluation, but that each school would be under an obligation to do some kind of school-based evaluation. In what follows I will give an account of why most teachers in Norway didn’t follow up this dictum, even if the political pressure was high and there were many different kinds of follow-up (Monsen, 1997).
I will start with an account of our experience in the project we started in 1989 and continued until 1995, and thereafter how it got a new direction with support from a Swedish project from 1995 (Granstrom & Lander, 1997). We were inspired by the action research carried out in England (Elliott, 1989) and in the project we tried to combine courses for teachers in school-based evaluation with research on how teachers experienced their first trials with this kind of evaluation as an integrated part of the course. Helen Simons, a well-known researcher in the field, was also involved in the project (Simons, 1987). Our first international presentation of this work, took place in a paper to AERA in 1993 (Indrebo, Monsen & Alvik, 1993). The year before we had organized an international seminar on school-based evaluation in Lillehammer (Alvik et al., 1992). Reading this report today, it can be characterised as “learning by doing,” where we believed it would only take the right combination of courses, practical experience for teachers and some kind of follow-up from knowledgeable persons outside the schools to do the trick. Most of the teachers and schools taking part in our courses, carried out school-based evaluation, even if for some of them this required great efforts and difficulties. They showed by their examples that it might be possible to do this on a bigger scale. In the paper I showed that in a national sample less than a quarter of the schools were doing any kind of school-based evaluation before the manual was issued.

From 1992 we shifted our strategy from courses on groups from schools to groups representing their municipalities, the reason was that we could then extend our reach beyond just a few schools. And these municipal groups were responsible for running courses for schools in their municipality and even try some follow-up counselling in individual schools. In the course period we still found the same thing: The people taking part in the course did their project with some kind of school-based evaluation inspired by the municipal level. In those first years we of course gained some insights into the complicated process of doing school-based evaluation. Both Alvik’s revised edition of his 1991 book in 1993 and Tiller’s book summing up experience from his project in six upper secondary schools (Tiller, 1993) and a few other articles from that period were still unshaken in their belief that when schools got the opportunity, they would do school-based evaluation.

SCHOOL-BASED EVALUATION AS PROJECT WORK

This simple and optimistic belief began to crumble in 1995 when we did the first follow-up study of what had happened in the schools and municipalities
after the courses had been completed. To my disappointment we could see that more than two thirds of the schools didn’t do any kind of school-based evaluation one to three years after the course had finished. The main reasons were that they had gained other important obligations (for instance, preparing for the new national curriculum) and that it was no longer a priority in their municipality. The last point is of special interest. It was confirmed by the people from the municipal level. Less than one third of the municipalities had school-based evaluation on their agenda, and for the same reasons as those on the school level. Municipalities in Norway are small, usually between 5000 and 25,000 inhabitants, and some taking part in our project were even smaller. With just a few people in the education office, they have to be very keen on what tasks to give priority. When the course was offered, it became a priority, because it was emphasized both on the national and the regional level. After the course was finished they felt they had done their duty, taken part in an important project, but as with all projects, they have an end point. They argued therefore that they had done their part and had to use their sparse resources on other and newer projects with a higher priority.

On the school level they understood this signal quite well, when it is not a priority on the municipal level, we don’t need to make it a priority either. The result was, therefore, reasonable from within the limits people both at the school and municipal level think they operate within. The question is, if it is possible to do something with these limits, and what kind of limits we are talking about. I will return to this question later on in this article. The conclusion being so far that even if the minister of education had said every school in Norway had to do school-based evaluation as a part of a new evaluation system, most of the municipalities didn’t take this literally. They felt free to make their own priorities, among other things, dropping school-based evaluation when they had done their duty in the project period.

A better understanding of why schools so often reject school-based evaluation, even when they know why they should do it, was revealed in a second round of follow-up studies starting in 1996. The previous year I had become involved in a Swedish research project about the relationship between schools and municipalities, the evaluation of schools and the municipality’s role in following-up. With funding from this project we could do a case study of two municipalities. The two municipalities we chose to study, were among the most active municipalities (in supporting their schools in school-based evaluation) and we knew them from our first study and from our course experience. The reason for choosing these municipalities was to find out what factors were of importance for schools doing and sustaining their evaluation activities. We soon found out that there was less school-based evaluation in the schools we interviewed than we thought should
have been the case with our knowledge of what was done on the municipal level. In both municipalities they had different kinds of supporting activities for many years, the first going back to a decision in the school board in 1989. They had supported and taken part in both our courses and courses from other colleges and a private research institute (IMTEC) and they had in different ways made a municipal organization of the school-based evaluation activities. In spite of these efforts, some of the schools in these municipalities did little of what can be characterized as school-based evaluation. Of course we found a greater number of more active schools than in the other municipalities we had earlier surveyed, but still more than half of the schools in both municipalities had no or very little activity in the year we carried out our interviews, even though some of them had done some “projects” as they called them and some were planning a new round of evaluation in the years to come. Most schools, also in these municipalities, confirmed the pattern we had found in our survey: doing projects in school-based evaluation were carried out in a sporadic manner, mostly to comply with pressure from above.

In the interviews we can see three sorts of attitudes towards the expectations of doing school-based evaluation. People who believed it was an important and necessary activity that should be done on a regular basis. In most schools their number was rather small and consisted mostly of principals and other teachers who took part in some of the courses arranged outside the school and who later took part in planning groups for school-based evaluations in their schools. Most of the teachers who composed the second group, had a rather pragmatic view of why they should do school-based evaluation: because it was an imposed demand from above and because they thought it would be good for the image of their school in the local community. For them the ideal solution was a project now and then to show they were capable of doing it. They didn’t see school-based evaluation as a natural part of their teaching role and as a necessary activity for school development (Scheerens et al., 1998). The third group of answers consisted of teachers who rejected the whole idea of school-based evaluation. In the schools we interviewed they were rather few in number, even if there was some variation between the schools. The difference between schools was related to role this opposition had in the total teacher group. Both in the pragmatic and oppositional group we found many teachers who had little knowledge about what school-based evaluation was supposed to be and why their schools were involved. Interpreting their answers it seems, they had a superficial relationship to the activity and they found it convenient if some of their colleagues could do most of the work. It is important to stress that in some of the schools these teachers were working, they had carried out some kind of school-based evaluation for several years. We must therefore conclude
that it might be possible for a school to concentrate this activity in a certain part of the teacher group and let the others be more or less spectators, and all these schools would be situated in municipalities with a more active involvement in school-based evaluation than in the other municipalities in the region.

**SCHOOL-BASED EVALUATION – DO TEACHERS FIND IT USEFUL?**

So far in this account of school-based evaluation in Norway, I have tried to show that even with municipal support, it is difficult for schools to integrate school-based evaluation into their repertoire of natural activities for teacher work. We have seen that most schools with some support and pressure can do it, but mostly as random projects. How is it possible to take the next step and integrate school-based evaluation into schools and the daily work of teachers? Before I can answer this question fully it will be necessary to complete some of the other research projects I have been engaged in and take a broader approach to the question. I would, however, start with a proposition: Most teachers do some kind of school-based evaluation or activities quite like it, but they do not think of it in this way. They all evaluate or assess their pupils, of course, and most of them make some comparisons between different pupil groups in the classroom, between different classes they teach and have taught. All these comparisons make it possible for most teachers to develop some kind of standards for pupil achievement and in Norway these standards are made more objective both by a system of exams and the provision of standardized tests in some subjects which the teachers can use if they want to. At the lower secondary level where exams are a part of the system, most teachers use them to form a baseline for their marks. It can therefore be said that pupil assessment is a core activity in the work of teachers and they regard it as essential and important. Why is it then so difficult to take the next step and use these pupil assessments in a school evaluation? A simple answer is that marking and exams do not necessarily tell us much about the quality of teaching and the school. As educational research documented many years ago, pupil achievement is influenced by many factors and teaching styles and teacher personality may have less of a role than is usually thought. Teachers know this and are therefore sceptical about marks and exams as indicators of school quality, as the educational boards in some counties have now published results from exams as an indication of quality. But what teachers have given little consideration to is the fact that their long experience of pupil assessment could be used for other purposes.
I have seen an example of this in another evaluation project I conducted for the county where our college is located. Some years ago all schools in this county had to make individual curricula for pupils with special needs. To help teachers write these curricula when their schools applied for extra resources for pupils with special needs, it was decided that they should use a standardized formula. My job was to find out how teachers used this formula and to thereby gain an understanding of the process (Monsen, 1995). The teachers' willingness to do the new and extra work entailed in writing these curricula might be seen in contrast to their willingness to do school-based evaluation. In writing these curricula, teachers had to observe pupils for some weeks, even months, assess their capabilities in several areas and specify these assessments and write them down for each pupil to find out how the national curriculum could be tailored to their needs. The first time they made these curricula, it involved a lot of work for them, and they had to learn a new way of assessing their pupils. Most of the teachers found this new obligation both challenging and useful and many saw the possibility of using these ways of assessing pupils and making curricula for local school development. The point of this story is not to refute explanations given by teachers or researchers as to why it is so difficult to do school-based evaluation, but provide a supplement and another twist. Teaching is an activity with much pressure on time; doing new things demands doing less of usual thing and as a consequence of this, teachers won't do new things if they are unsure of their usefulness. As can be seen from this example, teachers are willing to do both new and demanding tasks if they find them useful in their work. The problem with school-based evaluation compared with this example is that the majority of teachers are not convinced of its usefulness for their teaching, this despite some years with support and pressure both from the municipal and county level. As we have seen from both our own research and other research on this topic in Norway, the majority of teachers still seem to think that school-based evaluation is not something they desire or see as useful, instead they view it as an activity their superiors pressurise them into for political reasons. Even if it has been stressed several times from the hand book and onwards that school-based evaluation should be controlled by the school and that the school should own the results from the evaluation, we still find those who believe that all this talk about “school-based evaluation” is a cover for other more deeply felt intentions such as the quality control of schools for politicians and bureaucrats. Such beliefs have their history and I won’t go into it here, other than noting that it is related to the development that the role of teachers has been downgraded both economically and socially (Elliott, 1996).
TEACHERS CAN AND WILL LEARN NEW SKILLS

As can be seen from this example, possibilities exist for getting out of this seemingly stalemate position and as I said earlier we can start with teachers' experience with assessing pupils. I have another example for how this can be done from another evaluation project I have been involved in, where I and several other researchers have been evaluating a reform of upper secondary schools in Norway (Reform-94) (Monsen, 1998). As part of this reform the ministry issued a sort of a booklet of report cards called "The training manual." It was expected that every pupil should go through this training manual and fill out the report cards in cooperation with their teachers, the idea being that the manual documented the different curriculum activities each pupil had taken part in. This led to a storm of protests from teachers and the teachers' organization, who managed to stop the manual from being obligatory for all teachers. In my evaluation I found that a great majority (as could be expected) of teachers saw this as a new example of control from the ministry. But more interesting, and not so expected, the majority found some useful activities in the manual. One of them was (and still is) half an hour of a structured talk two times a year for each pupil with at least one of their teachers. In this individual talk the pupil has an opportunity of speaking about how (s)he was coping with their pupil role and more generally with their daily life in the school, and the teacher could give some feedback on how they (and other teachers in the class) looked upon the pupil and also give advice on how they might do better in school.

In many cases this developed into a mutual assessment from pupils and teachers where teachers obtained useful information about how pupils looked upon the school, the teaching and teachers. Most teachers appreciated the possibility of assessing of their pupils in a context, even if it was unnecessary for them to write down their assessment and only check out when and where the talk had taken place. And as in my other example, teachers also here had to do a new activity which they were not used to do and that was rather time consuming. The activity itself can be looked upon as a combination of assessment and counselling, and the counselling in particular is not a part of the teacher's role in upper secondary school, even if it is given an emphasis in the new curriculum. In a certain sense this talk can be looked upon as a starting point in a school-based evaluation if the teachers are able to use the information obtained, but most teachers didn't define this as a part of the school-based evaluation they were obliged to carry out (mandated in a ministry document from 1995). The reason was that they saw the information obtained as confidential and not for use in public. The most important point here is that most teachers had this talk (approx. 90%) and that they found it
useful. This contrasts with the mandated school-based evaluation in which teachers said they had taken part.

About half the teachers said they had taken part in some kind of systematic school-based evaluation (even though approximately three quarters of the teachers said they asked for and got oral information from their pupils about their teaching). Again the majority of the teachers said they didn't find this activity useful, but that they understood it as an obligation the principals had to meet from their superiors at the county level. Principals in some schools indicated in interviews that they tried to find ways of doing school-based evaluation without using this term because many teachers looked upon it as a provocation, and the activity as the interference of bureaucrats in their professional role in the "system." The possibility of doing school-based evaluation in upper secondary schools seems to be even more problematic than in primary schools, since teachers looked upon the activity as something they were pushed to do for bureaucratic reasons. But as we can see from my last example, teachers are willing to do new things if they find them worthwhile and useful.

How can it be possible to convince teachers that school-based evaluation may be an extension of activities in which they are already involved? In principle, there should be no big leap from using different kinds of pupil assessments to using these assessments in an evaluation of teaching and the school's curriculum. Our experience with the introduction of school-based evaluation in Norway from the early nineties has demonstrated the difference between theoretical principles and the implementation of those principles in the reality of the school. It is possible to convince a few teachers of the necessity and usefulness of doing school-based evaluation by working with them on courses where they can learn the theory and practice of this activity. But it is not possible to reach the majority of teachers by these rather intensive and time-consuming courses. Our two step organization using municipal groups with the responsibility for the following-up in the schools also had many shortcomings. I have therefore used two examples where teachers did new things with rather little support, a short introductory course and a combination of a clearcut method (the formula) and some pressure from administrators at the school level (head teachers and principals). But I think the main reason why these new activities were adopted so quickly and with such little resistance, can be found in the teacher's definition of their role as teachers. They thought of themselves first and foremost as cultural workers doing a professional job in the transmission of knowledge and skills to the next generation. To develop their skills and thereby doing a better job appealed to their professional identity and was therefore looked upon as a part of their responsibility. Most teachers took courses in the summer to upgrade their
skills and for some a course in school-based evaluation had professional appeal, but this was not the case for the majority.

THE PSYCHOLOGY OF THE ROLE OF THE TEACHER AND SCHOOL-BASED EVALUATION

Why do teachers regard school-based evaluation as a cover up for external school evaluation or as a bureaucratic activity of little usefulness for ordinary teachers? Most of them would admit when confronted with the question of the logical connection between the systematic collection of data and school development, that some school-based evaluation could be useful as part of their school-development activities. Their main objection is that it is too time consuming and seldom gives them the kind of feedback they need in the classroom. They can and do get this information by simpler and easier ways. They argue that too often they have experienced what is called school-based evaluation as a means of gathering information about the school as a whole, and while this can be useful in reports to municipal or county officers, it tells them little about pupil learning. At one level this argument explains why teachers use so much time on pupil assessment and why they are eager to learn new professional skills. On another level, with respect to the psychology of the teacher’s role, I would contend that another problem is involved. I find a logical gap between the teachers engagement in new methods of pupil assessment, for instance using standardized tests or one-to-one talks with pupils, and their refusal to engage in some kind of systematic data collection about pupil learning in the classroom where it might be possible to find some connections between teaching and learning in the classroom.

Most teachers are very anxious about the misuse of that kind of information. Therefore, we have seen that when teachers do school-based evaluation, only a minor part of it deals with the question of the relationship between teaching and learning and when it is done, they do not want this kind of information to get out of the school. There can be some good reasons for such anxiety; we are all familiar with what newspapers and politicians can do when they get hold of some of the information that is available. On the other hand, how can it be possible to do a better job as teachers without such knowledge? I would argue that this anxiety has deeper roots than a fear of misuse of information (Ball, 1997).

For many teachers digging into this relationship might lead to a loss of control over their teaching, others such as colleagues, pupils, principals and even parents taking over. Many teachers define their professional role in terms of their ability to maintain control over the whole teaching-learning process and fear too much interference into a vulnerable process (Blase, 1988). This is one reason why teachers find it so difficult to develop more collective responsibility for their
teaching (Little, 1996). There is a contradiction between teaching as a deeply personal relationship with your pupils and viewing teaching as a collective enterprise. They fear what they regard as investigations into this relationship, where what is investigated will be the more superficial and managerial part of the relationship, the personal way of doing things will not be appreciated (Gitlin & Margonis, 1995). In my evaluation research I have seen many examples of the ambivalence many teachers feel when it is said they should be more open to colleagues who observe each other, as well as evaluate and plan together.

Just one example can show how we can deal with this fear. In one of the schools where I have been engaged in case research, the new principal in the autumn 1994 when the reform was to be implemented, proposed that the whole school should take part in an investigation into the school as a learning organisation with a focus on teaching and learning in the classroom. This proposal was met with heavy resistance from teachers so the principal decided to postpone it and in the meantime convince teachers of the usefulness of the investigation (planned as a survey for teachers and pupils). It took nearly two years before he found the climate had changed and he could carry out the investigation. When it was, he had agreed with teachers that there should be some restrictions on what kind of information was to be accessible to different partners in the investigation. Information concerning individual teachers should be given only to the teachers concerned. The report should only differentiate between different departments in the school. Only the principal should have access to pupil evaluation of individual teachers and it was to be used only in individual counselling if the teachers asked for it, as many actually did, which says something about the trust this principal had built in those two years. Most of the teachers found this investigation useful and experienced the discussions it led to as an important professional challenge, it told them something about their school which they hadn’t known before. It was therefore decided to repeat this investigation on alternate years. Two important points are illustrated with this example; the first being that the teachers felt they had a way of dealing with the information to prevent misuse, and the other being how there was enough time reserved for convincing teachers of the usefulness of doing the necessary work and their cooperation in gathering interesting information, for instance on the teaching-learning process in the classroom (Haug, 1997).

**CONCLUSION: HOW DO WE CONVINCE TEACHERS OF THE USEFULNESS OF SCHOOL-BASED EVALUATION?**

I have seen in many other schools, how too much pressure from the outside, a lack of understanding of how vulnerable teachers are to this kind of intervention into
School-Based Evaluation in Norway

teachers personal spheres, rushing into the gathering data has created the kind of suspicion I have described earlier and this makes it difficult to develop the proper atmosphere for useful discussions of the information found. It should now be possible to put together some of the experiences and research done on school-based evaluation and achieve a better understanding of how to do school-based evaluation (Cousins & Simon, 1996).

(1) Before we make a national strategy for integrating school-based evaluation into a national system of evaluation, it is important to know about how teachers understand these new policies. In Norway, it was assumed that simple argumentation of why school-based evaluation was important and necessary for school development, presented in a manual given to all schools, would be taken literally. In the last ten years we have learned that convincing teachers about the usefulness of school-based evaluation is much more complicated. To convince them we must have a far better understanding of their resistance and not reject their arguments as purely a defensive protection of old privileges (Keltchermans, 1996). As I have tried to show with the two examples, teachers are willing to do new things if they can see them as an extension of their professional obligations.

(2) Convincing teachers that school-based evaluation can also be a useful and necessary activity for ordinary teachers, takes much more time than is available in an ordinary short course. In Norway believing what was said in the manual and in other official documents was rather problematic, taking into account what was said in the public debate about the poor quality in schools. Therefore, many of the demands for doing school-based evaluation which came from county and municipal levels were interpreted by the majority of teachers as a cover-up for external school-evaluation so popular in the public debate. Schools who succeeded in convincing their teachers, had a leadership with the ability to do it their own way and not give in to too much pressure from outside. They took their time and were able to demonstrate the necessary confidence and legitimacy. Then, as we saw in the last example, teachers were willing to set aside their suspicions and act as professionals and look upon evaluation as a regular activity for teachers.

(3) In many schools, school-based evaluation is done as a project with a beginning (often defined by a new course) and with clearly stated end (not so far from the end of the course). Many of the teachers taking part in the "project" look upon it as something they do because of their superiors and
because their school often gets some extra resources during the project period. Even if they find the experience interesting and stimulating, they still look upon it as something "extra," not done because they want to, and regard it as a natural part of their professional responsibility. It can be necessary to start school-based evaluation as a project, but the challenge is to transform it to an ordinary activity. This transformation should begin by looking at how teachers understand what they have been doing so far and finding ways of developing this into an extension of what they already do, in terms of pupil assessment.

(4) Support and pressure can be necessary both to start and sustain the activity, but what kind of support and what kind of pressure is crucial for the survival of school-based evaluation? Too much pressure from outside the school can be counter-productive. Most of the teachers will then be convinced that it is a cover-up. The pressure which provides the necessary energy comes from colleagues and principals expecting you to do a better job, from pupils, parents and other important groups in the school's local community who look upon the teacher as a professional. This is the kind of pressure where a good job can be done better, and not as can be seen in the national debates about education where teachers feel they are scapegoats for some national problems, such as failing in international competition. Support is also important, but as with pressure it can be counterproductive. We have seen local educational officers taking responsibility for collecting the data in the belief that it will stop the teachers complaining that they don't have time to do it themselves. The result has been that teachers look upon it as an external evaluation where they lack responsibility for data collection and therefore have no use for the findings either. The support needed should be like the principal’s in the last example: daily and informal contact where teachers can talk with the principal in small groups. In these "small talks" the teachers become clear about both the intentions behind and the usefulness of the evaluation. They have the opportunity to express their doubts, to try to work out their arguments in a relaxed atmosphere and even get some new ideas about how to use the new information in their own teaching. This principal had the necessary knowledge and skills to do this, but as we also have seen, many principals do not. Therefore, in every school doing school-based evaluation it is important to find the right person to take on the role. It might be an ordinary teacher, but usually (s)he will lack the necessary authority. It is therefore advisable to find someone in the school with the leadership qualities and authority to take this responsibility (Moller, 1996).
REFERENCES


Bredvold, R. (1997). *Co-operation between the Local Education Authority and the schools on Schools' Self-evaluation: A case study from two Norwegian municipalities*. Paper delivered at ECER, Frankfurt am Main (September 24-27).


EVALUATION AS SCHOOL IMPROVEMENT: A DEVELOPMENTAL PERSPECTIVE FROM ENGLAND

David Hopkins and Mel West

INTRODUCTION

The implicit link between evaluation and school improvement has been recognised ever since the term 'school improvement' became commonplace some 15–20 years ago, yet the exact relationship has only been intuitively defined. There are a variety of theoretical, methodological and practical reasons for making this relationship more articulate.

Theoretically the link between evaluation and school improvement is unclear and for those committed to both, that is an unsatisfactory situation. It is unsatisfactory not just because of a predilection for analytical tidiness, but because a lack of conceptual clarity about the field prevents one from taking control over its potential. We believe in the old adage that there is nothing so practical as good theory; consequently being unclear about the theory leads to impoverished practical action.

Methodologically, this clarity is needed because method is the link between theory and action. It informs both, as indeed they inform it also. Paulo Freire once commented that methodological confusion can always be traced to ideological error, and we are committed to avoiding that particular trap, particularly as the commitment to evaluation and school improvement is informed by a particular ideological position. The methodological implications
of this analysis form a major part of this chapter, as we describe in some detail the methodological activity contingent upon the analysis.

To be of any use to and in the field we need to be certain of the practical implications and applications of our method. This is particularly the case at the present time when evaluative activity is increasingly contractually regulated and consequently constrained by short time lines and shrinking budgets.

It is these philosophical, methodological and practical arguments that provide the rationale for the paper. Theoretical analysis is important because it implies different methodologies for each approach to evaluation and school improvement that, in turn, informs practical action. For reasons that will quickly become obvious, however, we are committed to the approach characterised "evaluation as school improvement." We explain this relationship by referring to a model that contrasts three distinct approaches to evaluation and school improvement. Following a brief overview of the model we subsequently use it to organise our analysis of recent changes in the English educational system. This critique provides further support for the "evaluation as school improvement" approach. We explore this in more detail by reference to our school improvement model "Improving the Quality of Education for All" (IQEA) using a number of methodological and practical examples. Before we launch into the argument however we should provide a brief definition of what we mean by evaluation and school improvement.

**DEFINING EVALUATION AND SCHOOL IMPROVEMENT**

Before we can begin our analysis of the three approaches to evaluation and school improvement we need to define our working definitions of these terms. Definitions of evaluation abound. Tyler (1949) originally defined it as 'the problem of determining to what extent educational objectives are being realised.' Following this lead most evaluators in recent years, particularly those working in North America, have as David Nevo (1986) point out:

Reached considerable consensus regarding the definition of evaluation as the assessment of merit or worth or as an activity comprised of both description and judgement.

Nevo (1986) continues to add, however, that:

A major exception to that consensus regarding the judgmental definition of evaluation is represented by the Stanford Evaluation Consortium group who defined evaluation as 'a systematic examination of events occurring in and consequent of a contemporary program - an examination conducted to assist in improving this program and other programs having the same general purpose' (Cronbach et al., p. 14). Cronbach and his associates (1980)
clearly reject the judgmental nature of evaluation advocating an approach that perceives the evaluator as ‘an educator [whose] success is to be judged by what others learn’ (p. 11) rather than a ‘referee [for] a basketball game’ (p. 18) who is hired to decide who is ‘right’ or ‘wrong.’

This last definition is consistent with our own approach that regards evaluation as an integral element in school improvement. We agree with Stenhouse (1975), when he argues ‘against the separation of developer and evaluator’ and in favour of integrated research. He continues:

> Evaluation should, as it were lead development and be integrated with it. Then the conceptual distinction between development and evaluation is destroyed and the two merge as research.

We perceive the fusion between evaluation and school development as defining the central axis of school improvement processes and roles. We are committed to school development and to the use of evaluative data to illuminate, inform and guide this process.

In using the term school improvement we imply a distinct approach to educational change that enhances student outcomes as well as strengthening the school’s capacity for managing change. School improvement is about raising student achievement through focusing on the teaching – learning process and the conditions that support it. It is about strategies for improving the schools’ capacity for providing quality education in times of change (Hopkins, 2000).

It follows from this that school improvement is about curriculum development, about altering the teaching/learning process, about sustaining changes in teaching and learning through changes in organisation and structures, and as such requires a developmental approach to evaluation. Such an approach to evaluation focuses attention on the process of strengthening the school’s capacity to deal with change and ensuring understanding of and commitment to the school improvement agenda.

With these definitions and the value positions they imply in mind we can now turn more fruitfully to our analysis of evaluation of, for and as school improvement. The differences between the three approaches are central to our understanding of the relationships between evaluation and school development.

### EVALUATION AND SCHOOL IMPROVEMENT – A TYPOLOGY

We have distinguished elsewhere (Holly & Hopkins, 1989; Hopkins, 1989), between three approaches to evaluation and their link with school improvement. These are evaluation of school improvement, evaluation for school improvement, and evaluation as school improvement. Such a distinction helps sharpen our
ability to work effectively in each of these areas as well as providing for a deeper understanding of the role of evaluation in the school improvement process. Ultimately, of course, each approach is equally important because in practice they build on each other. As we argue later, if evaluation of school improvement is done well this leads inevitably to evaluation for school improvement that in turn provides the substance for evaluation as school improvement (see Fig. 1). This is not to disguise however, the importance of 'evaluation as school improvement' as a mental-strategy for educational change and development.

'Evaluations of' are usually concerned with evaluation the outcomes of improvement effort or a particular initiative. These will tend to be built around the question 'what aspects of previous practice would we expect to see changed if this programme or innovation is to be effective?' Inevitably, therefore, 'evaluations of will tend to be 'product' evaluations, dealing primarily with quantitative and statistical data. Where 'process' evaluation is employed it, too, will tend to be a search for 'process outcomes.' Unfortunately, summative data produced as a result of evaluating school improvement activities have limited transferability to other innovations or to the distinctive cultures of other schools. It is inherent in our definition of school improvement that it is an 'adaptive, evolving and problem-coping' process (Louis & Miles, 1990) rather than one in which fidelity to other

Fig. 1. Evaluation and School Improvement.
Evaluation as School Improvement

initiatives – however successful – plays a large part. Consequently, ‘Evaluations of’ seem better suited to exercises arising from accountability motives than development. To foster school improvement, what is needed is an evaluation scheme that reflects the evolutionary, relatively autonomous nature of school development – and which is part of the development culture itself. This leads us forward into the next stage of the ‘cycle’ or ‘continuum.’

The sense of what is meant by ‘evaluation for’ school improvement is perhaps best captured in the commonly understood phrase ‘formative evaluation.’ This is evaluation conducted for the purpose of bringing about improvements in practice. The critical feature of this evaluation approach is that its prime focus is the facilitation of change. In this mode, though, the evaluation and the change remain distinct. There are two major issues raised in the literature that relate to the effectiveness of evaluation for school improvement. The first is its lack of integration with the knowledge about change and implementation processes. The second is the problem of communicating, utilising and gaining ownership of the data.

In order to do ‘evaluation for’ properly, Fullan (1983) argues that data would need to be gathered also on elements of the change process itself. To generate the kinds of information on which judgements and actions might be based, there would need to be three aspects of evaluation for school improvement:

(i) a definition of the change itself;
(ii) a description and assessment of the factors influencing change;
(iii) an evaluation of the outcomes of the change (both anticipated and unanticipated).

If ‘evaluation for’ is to facilitate improvement, Fullan and Park (1981) suggested that at a minimum the following dimensions would need to be incorporated into the evaluation design:

(i) alterations in structure:
(ii) use of new materials;
(iii) the acquisition of new knowledge;
(iv) the introduction of new teaching styles;
(v) the internalisation of new beliefs.

While we can accept that an evaluation design based on these implementation components would readily provide a framework for effective development our experience is that systematic tracking and integration of these strands is hard to achieve in practice.

Secondly, we have also found that the communication of evaluation data raises a further concern with the ‘evaluation for’ approach. The outcomes of
evaluation are usually presented in a written report or in statistical analyses. These can be particularly inappropriate methods for communicating about development, and the written form is a particularly difficult medium for moving the hearts and minds of school leaders and teachers. This communication problem has been shown to be significant (Cousins & Leithwood, 1986) and this is particularly the case when evaluation activities are conducted by external researchers.

The limitations of space preclude a full discussion of this issue, but it is probably sufficient to state that studies have shown (Cousins & Leithwood, 1986; Hopkins, 1989) that ‘evaluation’ activities can give rise to a great sense of ownership when they:

(i) focus on user issues;
(ii) involve users in the process;
(iii) develop a pedagogy for learning from evaluation;
(iv) make explicit the integration between evaluation and development.

This begins to take us beyond ‘evaluation for’ - and into the potential of evaluation activity itself as a developmental tool for user groups.

‘Evaluation as’ school improvement occurs when the evaluation has an explicit school improvement purpose and when those who will be involved in the development are also engaged in the evaluation. The process of evaluation and school improvement is, thus, one and the same thing, and there is no need to differentiate.

In the 1980s we saw the introduction of the GRIDS approach to school self evaluation (Abbott et al., 1988) in Britain, and other similar school-based review activities throughout other ‘western’ educational systems (Bollen & Hopkins, 1987). These can be seen as early examples of ‘evaluation as’ activities (Hopkins, 1988). Similarly, school self-evaluation, school development planning, action research and a variety of other school-based research activities are current examples of the integration between evaluation and development activities. Evaluation has the potential to be a change process in itself: to change teachers through their professional development; to change the culture of the school in which it takes place; to increase the knowledge base about the school; and to lead to further school improvement.

Based on our own research in schools and the research literature in this field, it appears that ‘evaluation as’ school improvement becomes increasingly effective when it:

- is ‘owned’ at the school level;
- is free from ‘accountability’ considerations;
• utilises a systematic and conscious methodology;
• relates to perceived needs at the school level;
• embraces a clear perception and definition of process and roles;
• is integrated into on-going school improvement processes;
• draws on external support when appropriate.

As noted earlier, when the model in Fig. 1 was first proposed (Hopkins, 1989), it was also argued that 'if evaluation of school improvement is done well, this leads inevitably to evaluation for school improvement, which in turn provides the substance for evaluation as school improvement.' However, our subsequent research work with schools both in the evaluative and developmental fields suggests that the reality is rather less straightforward. The 'cycle' implied in the model is in reality frequently interrupted. The movement from evaluation as; looking back; to evaluation as 'looking forward' is less sequential than we then suggested. The three approaches might be better represented as three overlapping but interdependent aspects of evaluation. The first providing descriptive or analytical data, the second representing our capacity to make appropriate judgements on which to design school improvement approaches, and the third representing a continuous process of data collection and school improvement response. We represent this modification to the model in Fig. 2. Our further engagement with the model within the context of the changing landscape of education in England

Fig. 2. Rethinking Evaluation and School Improvement.
during the past decade has also prompted us to rethink the application of the model to efforts for school-based development and reform. It is to an analysis of the relevance of the model to the changing educational context in England that we turn in the following section.

**THE CHANGING CONTEXT OF EVALUATION AND SCHOOL DEVELOPMENT IN ENGLAND**

In the previous section we described a broad classification of the different purposes served by and perspectives on the evaluation process. Although we still feel that this offers a useful overview of the evaluation process itself, it is clear that the educational reforms, which have been implemented in the U.K. over the past ten years, have altered the balance between these perspectives.

Whilst the previous decade saw much attention paid to the evaluation of initiatives intended to improve schools, the relevance of external, post hoc evaluation has been more difficult to discern in the newly established 'educational marketplace.' A number of factors contribute to this. One factor has been the change in central government policy with regard to systemic change. The use made of 'trailing' new ideas and approaches in a small number of educational authorities or schools has diminished sharply. Nevertheless many of the important developments of the 1970s and 1980s were introduced this way (e.g. Pupil Profiling, the Technical and Vocational Education Initiative, Schoolteacher Appraisal). Implementation limited to a small part of the school sector was evaluated, most often with a range of partners, and any lessons emerging could be assimilated before national implementation.

Generally, this approach to educational change was successful, and evaluation certainly played a part in the success. However, since the late 1980s the British Government has favoured a rather more rapid implementation strategy and change has tended to stem from legislation which affected all schools simultaneously. Subsequent evaluations have been commissioned, and some have led to modifications in policy, but the majority of schools may well feel that they were aware of the problems/difficulties associated with particular policies long before an 'evaluation report' became available. In that sense, the role played by systematic evaluation of policy implementation over the previous decade has been eroded, and such reports are now as much concerned with the demonstration of public accountability as the improvement of practice.

A second factor has been the demise of the local education authority (LEA) as a significant partner in the development of the educational system. Local management, with its transfer of funds and decision making into the school, has left the LEA with less influence and fewer resources. The district-level
evaluation of educational change has also declined sharply in this climate – LEAs do not have the funds to commission evaluations and, even if they had, schools might not be inclined to cooperate as they did in the past.

Third, there has been a growing dissatisfaction with the point of focus of 'externally' commissioned evaluations. In part, this relates to the concern of central government to establish that policies have been implemented, rather than that policies are achieving the purposes for which they were brought forward. Studies of implementation hold less interest for teachers than studies of impact. The dissatisfaction is also related to the failure (sometimes) to establish success criteria before implementation. This often results in 'evaluators' simply describing what they find, rather than monitoring for specific outcomes. In such circumstances evaluation is phenomenological rather than deliberate.

But, this reduction in the evaluation of the impact of school improvement initiatives has been offset by a widening of evaluation used for school improvement and evaluation as school improvement (see Fig. 3). Again, we can identify factors which have contributed to this development – the two principal ones being the new pressures on schools to demonstrate their own effectiveness, and the growing understanding of what local management means at school level.

---

**Fig. 3.** The Change in Balance of Evaluation in the United Kingdom over Recent Years.
The recent emphasis on school performance, fuelled by the production of a variety of ‘measures’—such as the league of tables of public examination results and the testing of all pupils using the standard attainment tests (SATS) tied to the National Curriculum, has made schools acutely aware that, in the educational marketplace, results count. Indeed, the climate of expectation that has been generated by this emphasis is such that the school may feel that even to maintain performance—however satisfactory the level—is not good enough. In a climate where continual improvement is looked for by parents, then to stand still will be to appear to go backwards. Consequently, the quest for continuing improvement brings with it a need for new measurement—schools need to know more about how they are doing, and where they might improve, so that evaluation for improvement is now often seen as a major priority.

The second factor is associated with the new pressures and opportunities that come with self-management. It is now increasingly recognised that the future of the school, its ability to develop (or, in some contexts, simply to survive) over time, will depend on the quality of decisions which school managers are making now, for the future. Insofar as evaluation is, above all, a process which seeks to influence the quality of decisions about the future (here we can distinguish between evaluation, with its commitment to influence future decisions, and research, which, though it may influence decisions, is concerned principally with explanations of the past) it is understandable that those responsible for decision making in the school are looking to evaluation as a source of knowledge and understanding.

The level of activity associated with evaluation for school improvement has increased markedly in the post Education Reform Act period. Two examples of this emphasis which involve all schools are the adoption of Development Planning, and the imposition of formal external inspection under the supervision of the Office for Standards in Education (OFSTED).

Development Planning (see Hargreaves et al., 1989; Hargreaves & Hopkins, 1991), which all schools now carry out annually, is a continuing cycle of review, planning, implementation and further review. This process helps schools to take stock of existing strengths and weaknesses, identify priorities for development, plan to review those priorities, and evaluate their success. In the majority of schools this cycle of action and review has become the dominant approach to planning, so that evaluating what needs to be done to improve the school is an activity to which most teachers will make some contribution. Unfortunately, schools seem to have mastered the review element more quickly than the planning or implementation elements, so that whilst most schools could offer a grounded picture of what needs to be done to move the school forward, it is less certain that they will have plans which will enable this to be done, or the
capacity to implement plans in a way which does move the school forward. Nevertheless, there is clear evidence within individual school’s development plans that schools now systematically evaluate their own performance and identify action points - a large step forward from the mid-eighties.

The school’s commitment to use evaluation to identify opportunities for development has been reinforced by inspection. The inspection process is clearly documented (see OFSTED, 1995) and identifies a range of areas which will be scrutinised and, for each, a set of criteria which will be applied in reaching judgements about the school’s performance. Following the inspection, a report is produced according to the format set out in the Handbook for Inspection. This report is a public document – it is available to the staff in the school, to parents, to the local community, and it requires a public response; the school’s Action Plan following inspection. Despite our reservations about the inspection process (Hopkins, West & Skinner, 1995) it is clear that it is seen as an important review of how the school is functioning, and that the ‘key issues for action’ are seen as areas where responsible action must be made. To this extent, inspection is clearly an important vehicle, conveying evaluation for school improvement onto the agenda of all schools. The more recent emphasis within OFSTED on using the framework as an aid for planning (OFSTED, 1994) and for school self evaluation (OFSTED, 1999) is further evidence of this trend.

The trend away from evaluation of towards evaluation for is, then, an important one. It transforms schools and teachers from objects of the evaluation process into partners within the evaluation process. But, as we hint at above, our own feeling is that the most exciting and, thus far, underdeveloped use of evaluation within schools lies in the third strand. It is evaluation as school improvement that offers schools the best opportunity to build a continuously developing culture. It is evaluation as school improvement that can best help the school constantly to improve its ability to meet the needs of students and to renew itself in the face of ever-increasing demands from government and expectations from parents. Though, as yet, only a minority of schools have embarked upon a systematic programme of evaluation and improvement, interest and activity is certainly growing. In the rest of the chapter we describe how we work with schools on their evaluation and improvement agendas.

**LOCATING EVALUATION WITHIN A MODEL FOR SCHOOL IMPROVEMENT**

We have been fortunate to work with some of those schools who are lighting the way forward. Over the past ten years, we have been able to test out our belief that evaluation is at the heart of school improvement through the
'Improving the Quality of Education for All' (IQEA) Project, which we developed with Mel Ainscow whilst we were all based at the University of Cambridge Institute of Education (Ainscow & Hopkins, 1992; Hopkins, 1994; Hopkins & West, 1994; Hopkins, Ainscow & West, 1994; Hopkins, West & Ainscow, 1996). Our approach can be outlined through reference to the 'model' of school improvement that underpins our thinking. We do not start with abstractions about the kinds of 'vision' or goals that may 'inspire' the school community, as we have ample evidence from our previous work with schools that quality of vision is independent of quality of schooling. Many schools with entirely laudable vision statements seem unable to reflect these in practice, and conversely, many of the best practices we have seen have never been formalised into school goals. Rather, we encourage 'IQEA schools' to begin by evaluating the quality of experience they currently offer their students – the quality they deliver in the classroom, not the quality they aspire to in plans.

Essentially, we are asking Project schools to start from where they are, rather than to imagine where they would prefer to be. We have found that when schools look closely at what they currently provide, and then, in the context of their own constraints and opportunities, consider what can (or indeed must) be done, the generation of goals and priorities for action remains ‘grounded’ in the realities of the school.

The conceptual model of how quality improvement takes place can thus be simply illustrated as in Fig. 4. Our assumption here is that the starting point for improvement effort is student outcomes – indeed, we might define improving the quality of schooling in these terms, as a deliberate programme aimed at reducing the discrepancy between the outcomes we desire for our students and the outcomes they actually achieve. The centrality of evaluation to such a model is obvious.

The desired improvements in these outcomes form the basis for specific goals – the priorities that guide and focus teacher energies. But we cannot expect ‘priorities’ to galvanise efforts if there are too many. Often, this means that decisions about priorities must be made, moving from the separate, perhaps even conflicting, priorities of individuals or groups to a systematically compiled set of priorities which represent the overall needs of a whole school community.

In their book on development planning, *The Empowered School*, Hargreaves and Hopkins (1991) have suggested that two principles should guide this process of choice among priorities:

(i) Manageability: how much can we realistically hope to achieve?
(ii) Coherence: is there a sequence that will ease implementation?
To those principles we have added a third,

(iii) Consonance: the extent to which internally identified priorities coincide or overlap with external pressures for reform.

In practical terms, this means that within the ‘family’ of IQEA project schools each school is pursuing its own particular improvement priority. The IQEA project offers a way of thinking about and working on school improvement, but the decision about what to improve must in each case be determined by the individual school, arising from its evaluation of individual circumstances and opportunities. We have previously written (Hopkins, Ainscow & West, 1994; Hopkins, West & Ainscow, 1996) about the range of priorities project schools have selected, in an attempt to give some sense to the sorts of activity going on. Such accounts explain why the notion that each school should focus on priorities relevant to its own particular circumstances is central to our conceptualisation of the improvement process.

They also illustrate how ‘capacity’ can be created and tapped into to support the school’s improvement work, since it will be the quality of the school’s management arrangements and their classroom arrangement that determine
whether the priorities identified lead to improvements in outcomes or not. Essentially then, the ‘arrangements’ are mediating variables, through which ideas about improvement are given substance. But, all our experience tells us that establishing these arrangements in the school will have limited impact unless teachers can understand what is being done and are then able to associate the development of management or classroom practice with tangible goals. Hence the need for emancipatory evaluation, both to determine the priorities, and subsequently, to investigate how the priorities can best be tackled. Of course, once a particular condition has been established it may well mean that other goals will be easier to pursue – this is why we describe this process as ‘capacity-building,’ since it creates the sorts of infrastructures and develops the sorts of skills (at both school and classroom levels) that facilitate the implementation and evaluation of change of whatever kind.

The strategy followed in the project is, then, relatively straightforward, at least conceptually. Identify areas for improvement from an evaluation of what is currently happening, select a limited (we discourage schools from trying to pursue more than two or three priorities at one time) number for action, then develop the school’s management and classroom arrangements as necessary to enable the priority to be pursued in appropriate and supportive conditions. We believe that ‘IQEA schools’ have demonstrated the wisdom of linking activity related to priorities with activity to develop the school’s management and classroom arrangements. We have also noted that this process becomes ‘easier’ with time, as the schools develop capacity and learn to use it, building successive improvement in successive years.

We have suggested (Hopkins, Ainscow & West, 1994) that by addressing specific improvements alongside a more general commitment to ensure that the best possible ‘arrangements’ are in place, the school is, in effect, developing the culture, rather than simply organising or re-organising around current priorities. Indeed, we feel that an effective school improvement strategy offers the most reliable means of enhancing school culture. Most often, this process of systematic internal evaluation has been approached through an initial analysis of outcomes currently generated for students. However, one of the key management arrangements we associate with the development of ‘capacity’ is ‘Enquiry and Reflection’ (see Ainscow et al., 1994, Chap. 3; Hopkins & West, 1994) and one of the key classroom arrangements is ‘Reflection on Teaching’ (see Hopkins et al., 1997, Chap. 8; Hopkins et al., 1998).

It is through the ongoing development and spreading of the habits of enquiry and reflection amongst the staff that continuous evaluation of the school’s impact and response to the issues this raises is achieved. In this sense, within the IQEA Project we have promoted ‘evaluation as’ improvement through the
development and application of enquiry skills in all aspects of the school’s operations – management or teacher focused.

**EVALUATION AS SCHOOL IMPROVEMENT: TWO METHODOLOGICAL EXAMPLES**

Before concluding the paper we thought that it would be helpful to provide more concrete examples of the way in which we work with schools in order to build their capacity for ‘evaluation as school improvement.’ Given limitations of space we have selected two contrasting examples that give some indication of the range of activities with which we engage.

*Mapping the ‘Process of Change’*

During our programme of research and development into school improvement, we have come to recognise that traditional methods such as evaluation interviews, questionnaires and observations are sometimes too cumbersome and time-consuming in disclosing the intricacies of the change process. Moreover many teachers are ‘over-evaluated’ by these standard techniques and in consequence are sometimes reluctant to find time for yet another interview or questionnaire. There is room for new, more user-friendly yet penetrating, techniques for investigating and measuring the complex processes and relationships involved in school change. It is because we found only limited methodological guidance in the existing research literature that we sought a grant from the ESRC to remedy this important lacuna. The grant gave us the opportunity to develop six new techniques for mapping the process of change in schools. A comprehensive description, with advice on administration, of the techniques is found in the manual produced as a result of the research *Mapping Change in Schools: The Cambridge Manual of Research Techniques* (Cambridge University, 1994).

The six techniques in the series cluster around two key elements in the change process: the individual teacher and the school as an institution. Despite the proliferation of externally mandated changes, the success of many change initiatives remains attributable to the commitment of individual teachers. Certainly the impact of any change on student outcomes is heavily affected by behaviour in the classroom. At the same time, it is claimed that the school as a whole, especially its climate/ethos/culture, makes an important contribution to development and change. Three of the techniques focus on the *individual teacher* and tap data at that level. The other three focus on the *school as an institution*. Data from both levels are essential if the interaction between
individual and institution in processes of change is to be better understood. (This brief description of the mapping techniques is adapted from Ainscow et al., 1995.)

The manual of techniques is accordingly divided into two series:

**Series 1: Individual (teacher) level**

- Technique 1: The Time Line of Change
- Technique 2: The Experience of Change
- Technique 3: The Initiation of Change

**Series 2: Institutional (school) level**

- Technique 4: The Culture of School
- Technique 5: The Structures of School
- Technique 6: The Conditions of School.

*Technique 1: The Time Line of Change*

The aim of this technique is to record how individuals within a school perceive their experience of a particular change over a period of time. In order to stimulate individuals in recollecting their involvement within what occurred, a timeline diagram is prepared which includes a small number of key events associated with the change. This is presented to the respondents who are invited to annotate it, noting additional events that seem significant to them. They then comment about the levels of involvement they felt during the period of time represented. These annotated timelines are collated and analysed in preparation for a semi-structured interview with each respondent. Data from these interviews are then collected to provide a composite historical perspective on the change. This provides a good starting point and framework for using the other techniques.

*Technique 2: The Experience of Change*

The purpose of this technique is to gather information about the feelings of individuals towards changes in their school. It involves a 15-minute activity during which the respondent is encouraged to explore their feelings about a recent change of major event. The respondent is presented with a set of words printed on individual cards and asked to sort them into a series of piles with respect to their feelings about this change. The feelings that are seen as being most significant by the respondent or the basis of a more in-depth discussion. This allows the individual to give their own account of their choice of words. The technique lends itself to use in conjunction with Technique 1, The Time Line of Change.
Technique 3: The Initiation of Change

This technique taps teachers' commitment to change and their sense of control over it. It differs from the previous two techniques in that it is concerned with change in general rather than a specific change. The technique also maps the contrast between teachers' commitment to, and control over, changes initiated within and outside of the school. Two sets of five quotations from teachers for both internal and external change are presented in a random order, and participants make the one that most closely accords with their own views. There is also an opportunity for teachers to add a comment of their own. The responses are easy to quantify, and can be rapidly and graphically fed back to groups of staff. The contrast between control and commitment is often vivid, as in the difference between school as opposed to government-initiated change. Feedback can lead to intense debate about teachers' attitudes towards internal as opposed to external change, and about the nature of commitment and control, a relationship which is rarely discussed in school.

Technique 4: The Culture of the School

The purpose of this technique is to generate data on teachers' perceptions of the culture of their school, the direction in which the culture is moving and their ideal culture. The technique is presented in the form of a board game, with four 'players' (Hargreaves, 1995). A square board, in the form of an 8 x 8 grid with differently coloured corners, is laid on the table. Each player is given a set of four cards, each in a different colour to match the corners on the board. On each card is a description by a teacher of a school culture. Players have a personal copy of the grid and mark a square where their own school's culture is located. From this point an arrow is inserted to indicate the direction in which the school is moving. Another mark is placed where players would like the school culture to be. The four players then share their individual responses. Using the board on the table, they discuss the issues with a view to reaching a group consensus on the three required responses.

Technique 5: The Structures of the School

The purpose of this technique is to generate data on some of the basic social structures underlying school cultures. Five structures are posited: political (the distribution of power, status and resources); micro political (the manoeuvrings of interest groups and alliances); maintenance (routines and procedures to ensure orderliness and continuity); development (new structures to accomplish innovation and change); and service school cultures, called 'traditional' and 'collegial.' Participants are presented with two teacher-written multi-dimensional cameos describing each type of structure, one of a strongly
traditional and one of a strongly collegial school. Teachers locate their school on a scale between the two and then are given an opportunity to describe or comment in writing on their own situation. Finally, teachers suggest an ideal position for their school. By collating the individual responses, a map of the school’s perceived cultures, actual and ideal, is obtained.

**Technique 6: The Conditions of School**

This technique consists of a scale for measuring a school’s internal conditions and potential for innovation. The 24 items are grouped under six headings which represent the key conditions necessary for school improvement. These six conditions comprise the various aspects of a school’s development structure. The literature suggests that these conditions are the main determinants of a school’s capacity for managing and supporting change. The scale can be used as a diagnostic instrument for identifying weaknesses in the school’s development structure, as a means of measuring a school’s progress in creating and modifying its development structure over time, and for comparing the ‘change capacity’ of different schools. The scale is easy to administer and analyse. It is also amenable to further analysis by subgroup. The technique also complements and could cross-validate other techniques in the manual. When used as part of a school improvement strategy it gives staff a more sophisticated language with which to discuss and plan school development.

In the same way as there are various ways of interpreting the data generated by the techniques, so there are different ways of using them for ‘evaluation as school improvement.’ We have developed three different approaches to feeding data back to schools.

First, because the techniques are amenable to rapid analysis and presentation, virtually immediate feedback can be given in some cases. This is especially true of the culture game, school structures and responses to change techniques. When used with teachers as part of a training day or staff meeting, we found it possible to give aggregate responses to the whole staff during the same session. This feedback often aroused great interest and animated discussion.

The second method of feedback that we used was to talk through the results from the administration of the whole battery with the senior staff of the school involved. As we had developed a method for reducing the data onto one side of A4, this made the conveying of the information much easier to handle. Our approach was to present the data in a sequential and descriptive way, and then on the basis of this to encourage discussion and interpretation of the results from the senior staff. This proved at times to be a delicate process. Occasionally our ethic of confidentiality was challenged and we were often encouraged to be judgmental rather than descriptive and so go beyond the data.
The third approach to giving feedback is where the data are used as part of a school improvement process within a school. We have begun to take individual aspects of the data in some of our IQEA schools as a basis for action planning. The experience we have had convinces us that the data emerging from the techniques has great heuristic power.

To summarise, when taken together the techniques in the series provide a map of the process of change in a school. They can be used individually to investigate particular aspects of change processes or in combination for a more comprehensive analysis. The techniques can also be used to map changes in the school's conditions over time and facilitates the process of change and improvement in schools.

**Linking Evaluation to Instruction**

IQEA in common with a number of school improvement projects throughout the world, has sought to develop the capacity of schools to accommodate and use external change in order to maximise student outcomes. As we have seen, the initial work based at Cambridge, sought to develop the management conditions of the Project schools. Drawing from the already extensive literature on effective teaching and the burgeoning school effectiveness literature, subsequent school-based research has suggested to us that there is also a set of classroom conditions which also needed to be identified in order to establish the foundation for effective teaching and learning (Hopkins et al., 1997, 1998).

One of the conditions identified has been the teaching repertoire of staff (Hopkins et al., 1997, p. 60):

> Effective teachers have a range of learning activities, tasks or experiences for pupils which they know are successful in bringing about certain types of pupil learning... In practice, more effective practitioners have a range of teaching skills, styles, models and approaches which comprise a teaching repertoire.

Sharing the belief that 'models of teaching are in fact models of learning' (Joyce et al., 1997, p. 8), we believe that students need to be conversant in the learning techniques required by the teaching methods employed in lessons in order to take best advantage. Stated simply, students who know how to work effectively in groups will learn better when teachers organise group work than those who do not have the necessary skills. For an extension of teachers' repertoire to be most effective, there needs to be a concomitant extension of their students' learning repertoire.

There is already an extensive literature on the component parts of effective teaching (see, for example, Hopkins, Ainscow & West, 1994, Chap. 4) but less on the process of matching teaching strategies to students' learning styles. Much
of the matching of teaching and learning styles has been extremely speculative, based upon the premise that if a sufficient variety of strategies is employed, then a catch-all effect will apply; for example the phrase ‘Increasing the range of learning experiences provided in our schools increases the likelihood of more students becoming more adept learners.’

The need for some form of dialogue between teachers and students about teaching and learning methods in the classroom has increasingly been recognised by a number of the schools in the IQEA project. These schools have shown themselves willing to interrogate students on their views about what constitutes effective teaching. It is also clear that they regard some acknowledgement of student learning preferences, in the teaching which takes place within their classrooms as an element of effective teaching in its own right. They have also called for an easy to administer research instrument that can both help them match what goes on in classrooms more closely to the preferences of their students and provide clues about where to develop the teaching repertoire of their teachers and the learning repertoire of their students.

In order to undertake an audit of the teaching strategies used in its classrooms, and a survey of students’ views on those strategies, we developed instruments based on the work of David Kolb. Kolb’s seminal work, *Experiential Learning* (Kolb, 1984) effectively reconceptualises Piaget’s work on developmental learning into four distinct and authentic learning styles, with no implicit hierarchical structure. These four learning styles can be represented as quadrants in a grid where the two dimensions of perceiving and processing information have been juxtaposed, and Kolb also gives useful descriptors of each learning style.

Our colleagues further identified a range of classroom activities and strategies associated with each of the four learning styles (see Fielding, 1994) and from this produced an observation schedule which could be used to record the incidence of these various activities in a lesson (Beresford, 1998). Each activity is coded according to the learning style for which it caters. As each activity occurs in the lesson, its incidence is noted. There is no assessment attempted regarding the effectiveness of the various strategies within the context of the lesson. At the end of the period of observation the different number of strategies and learning activities employed by the teacher is totted up and recorded, in the boxes provided, against the appropriate learning style. Hence the lesson can be said to have a particular profile corresponding to the combination of numbers in the boxes. These can be converted into percentages of the total number of strategies and activities used.

In order to assess students’ preferences for these characteristic teaching activities, our colleagues drew up a similar schedule on which students were
asked to indicate which of the activities they preferred. The schedule consists of a list of classroom activities directly related to the teaching strategies listed in the observation schedule. By scoring 'Don’t Like' responses as 0, 'Don’t Mind' as 1 and 'Like' as 2 and adding the total for each of the learning style categories, a profile similar to that derived from lesson observations can be derived for each student. By adding the totals of all students in a particular group, a group profile can be obtained. These profiles indicate individual and group learning style preferences.

The schedule is versatile inasmuch as it can be used to gauge individual’s learning preferences as well as group ones. Students’ preferences in individual subjects can be assessed as well as their general learning preferences. A minority of schools have used the schedules to find out which strategies the students feel are most effective in the teaching of an individual subject but most have felt that their students lack the necessary analytical skills to arrive at such a judgement. The schedule can also be used to assess any gender differences or differences between year groups.

**EVALUATION AS SCHOOL IMPROVEMENT: BUILDING THE CAPACITY FOR ENQUIRY AND REFLECTION**

We have previously (Ainscow et al., 1994) set out the case for enquiry-driven improvement efforts:

We have observed that those schools which recognise that enquiry and reflection are important processes in school improvement find it easier to sustain improvement effort around established priorities, and are better placed to monitor the extent to which policies actually deliver the intended outcomes for pupils. Central to the conditions that promote the effective use of enquiry and reflection as developmental tools are:

- Systematic collection, interpretation and use of school-generated data in decision making.
- Effective strategies for reviewing the progress and impact of school policies and initiatives.
- Widespread involvement of staff in the processes of data collection and analysis.
- Clear groundless for the collection, control and use of school-based data.

Whether initiated externally to facilitate judgements about a school’s performance (e.g. inspection), or internally to review the progress of policies or programmes (e.g. mapping techniques, or teaching and learning inventories), we have found that schools are increasingly interested in the generation of information which can be used to assess how the school is ‘working.’ One of
the ironies we associate with this interest is the fact that information gathered by outsiders, be they inspectors or consultants, is often seen as having more significance than information which is routinely available to those within the school community. But, we have also observed that where schools understand the potential of internally generated information about progress or difficulties, they are better placed to exploit opportunities and to overcome problems.

We believe, therefore, that in a large number of schools the range of data available is being under-used. Instead of being brought to bear on decisions confronting the school or being used to increase understanding of present strengths and weaknesses, information is overlooked, and uncoordinated. A further irony is that often considerable effort has been put into the amassing of such information, as we frequently record, report or file pieces of information despite there being little prospect of purposeful use.

Of course, some schools are much better organised in this area, and have clear systems and procedures for collecting, analysing and interpreting information which is seen as relevant to particular aspects of the school or particular decisions. Even in these cases, however, a more general commitment to enquire into and reflect on the school’s progress is rare – more often it is the issue that is identified than the information collected, rather than data being collected to help identify what the issue should be. We would not want to suggest here that everything that takes place in a school can be noted, nor that all information has equal significance. But our work with schools that have adopted a sustained commitment to improvement initiatives however has led us to identify the habits of enquiry and reflection as important forces for improvement. This is the basis of ‘evaluation as school improvement.’

CODA

We have argued in this paper that evaluation can function as an improvement strategy when the conditions in school are appropriate. It has stated that it is not evaluation of schooling per se that matters, but the potential of evaluation as a development tool that commends it to us. Consequently we believe that evaluation only becomes a worthwhile activity when it brings with it a commitment to action, some kind of automatic response which is seamlessly joined to evaluation processes. Our experience within the IQEA Project suggests that it is possible to make this link, to join reflection to action, to graft improvement onto the rootstock of evaluation. However, it requires a clear commitment within the school community, no little energy and, above all, the habit of enquiry amongst teachers.
ACKNOWLEDGMENTS

The present chapter reflects on some fifteen years of action research that has focused on the links between evaluation design and school-based development. As a result it inevitably draws on some already published material often written in collaboration with colleagues; in particular Holly and Hopkins, 1988; Hopkins, 1989; Hopkins et al., 1997; West and Hopkins, 1997. All other previously published work used in the chapter is specifically referenced. We are grateful to our colleagues, in particular Mel Ainscow, John Beresford, David Hargreaves, Peter Holly and David Jackson for their clarity of thought, intellectual generosity, and enthusiastic commitment to collaboration.

REFERENCES


School Self-Evaluation in the Netherlands

Jaap Scheerens and Maria Hendriks

Introduction

In this chapter the policy-context of school self-evaluation in the Netherlands is described as a basis for the “let many flowers bloom” orientation in the field of school self-evaluation. Traditionally schools have had considerable autonomy in the Netherlands. More recent policy initiatives have even increased school autonomy, particularly in the domain of educational finance. In the curricular domain there is a gradual move in the opposite, more centralized direction; at least in primary and lower secondary education. This is manifested by stating (still general) attainment targets and by strengthening accountability oriented evaluation and assessment forms.

The very recent “Quality Law” offers a framework that encourages schools to carry out school self-evaluation, without providing detailed formats on how to go about this. In this context, where a strongly developed public educational support structure is a significant factor, there is a broad supply of approaches and tools for school self-evaluation.
In the Netherlands, schools have been free in the choice of the principles on which they are based and in the way they organize their teaching (freedom of foundation, conviction and organization, Ministry of Education, Culture & Sciences, 1999, Van Oijen in Solomon, 1998). This freedom, guaranteed under article 23 of the Constitution, results in a variety of school-types: public schools, denominational government-dependent schools and schools that are based on a particular pedagogic philosophy like Montessori schools, Dalton schools and so-called “free” schools based on the philosophy of Rudolf Steiner. Groups have the right and possibility to found schools, based on religion, philosophy of life or pedagogic-didactic ideas. In addition, within the global standards of “soundness” set by the Ministry, schools are free to determine what is taught and how (the objectives, the subject matter, the teaching materials, the evaluation methods and the allocation of the instruction time per subject). These standards, which apply to both public and private education, prescribe among others:

- the range of subjects to be studied;
- the general attainment targets or examination syllabuses and the content of national examinations;
- the number of teaching periods per year, the qualifications which teachers are required to have;
- the participation in decision making of parents and pupils in school matters;
- the planning and reporting obligations (Ministry of Education, Culture & Sciences, 1998).

The traditional autonomy in the pedagogical domain, before long, was not matched by a similar degree of freedom in the domain of educational finance and the conditions of labor of the teaching force. However, from the eighties onwards, further decentralization has been initiated. In 1993, in the so-called Scheveningen Talks on Administrative Reform, the Ministry of Education and Sciences and the organizations of governing bodies of the schools agreed upon more autonomy for the schools in the domains of administration and finance, and on decentralization of some tasks to the municipalities. To some extent schools were also given specific responsibilities to take care of quality evaluation, although at the same time more centralistic forms of quality care were enforced.

In connection with this shift in educational policy the role of both external and internal evaluation changed. Since the legal requirements for the soundness of education have become less detailed and prescriptive, external evaluation by
School Self-Evaluation in the Netherlands

the Inspectorate shifted from a legalistic supervision to a more substantive educational monitoring, based upon evaluation criteria and standards developed by the Inspectorate itself.

So, the fact that schools, to an increasing degree, are expected to bear responsibility for safeguarding the quality of education is to be seen as a direct consequence of the government policy of decentralization and deregulation. At the same time there is a growing interest among the ‘clients’ of the school to know about school functioning, with respect to both processes and outcomes. Parents and pupils are more and more critical of the quality of the education offered. The institutes for secondary, further and higher education have also become more explicit in stating what they expect from lower forms of education. There are also signs that the society at large – employers, interest groups and organizations concerned – are becoming increasingly interested in education and the quality that is actually achieved.

As it was already hinted at in the above, decentralization has been counter-balanced by a few more centralistic tendencies in the substantive educational domain, particularly in primary and lower secondary education. These shifts are evident in a stricter delineation of the timetable (range of obligatory subjects), a tendency to further specification of attainment targets, for example by presenting “intermediary educational objectives” for primary schools and also in external school evaluation. Decentralization and centralization tendencies concerning school evaluation will be further explained by referring to the “Quality Act” of 1998 and its implications for external and internal school evaluation.

The ‘Quality Act’

At August 1, 1998 the law concerning the quality policy (the ‘Quality Act’) went into effect. In this law it is laid down that schools themselves are responsible for the quality of education they provide and the pursuance of a quality policy to ensure its improvement. It has also been decreed that all schools develop a system of quality assurance.

In the law the agreements made in the ‘Scheveningen Talks on Administrative Reform’ in which the authorities and the organizations of governing bodies of the schools agreed upon a development to greater autonomy of schools and accompanying processes of consultation and accountability, are elaborated. As announced in the policy document ‘The school as a learning organization: quality policy at primary and secondary schools’ (Ministry of Education, Culture and Sciences, 1995), the school is required to develop three policy documents: a school plan, a school prospectus and an arrangement for complaints.
School Plan, School Prospectus and Arrangement of Complaints

The school plan outlines the schools' policy on the quality of education, including its policy on educational matters, staffing and internal quality assurance. These are the only three compulsory elements. Other topics may be incorporated at the school’s discretion. The school plan can be regarded as an integral policy document, but also has the function of an accountability document to the Inspectorate. Schools had to send their first school plan to the Inspectorate before August 1, 1999. The school plan has to be updated every four years.

The school prospectus contains information on the schools' objectives, the educational activities and the results achieved. It is a public accounting of schools to parents and pupils, giving them the opportunity to enter a dialogue with the school, enabling them to compare schools and facilitating them to choose the most suitable school. As from 1 January 1999 schools are obliged to have a school prospectus. The school prospectus has to be updated every year.

The arrangement for complaints offers a supplement to existing possibilities of involvement and participation of decision-making of parents, pupils and staff. The intention is to lower the threshold for parents for registering complaints about the school boards and members of the staff. In this way, the schools themselves receive signals, which could help them to improve their education and overall functioning. As from 1 August 1998 at all schools an arrangement for complaints has to be available (Ministry of Education, Culture and Sciences, 1998; VVO 1998a, b; WVO, 1998).

In addition, from 1996 onwards the Ministry of Education, Culture and Sciences publishes an annual national guide to primary and secondary education with the aim to inform parents and pupils about their main rights and obligations vis-à-vis the school. The guide is also intended to help parents in choosing a school for their child.

The school plan, school guide and arrangement for complaints have to be regarded as means to an end. Together with other means, they could stimulate more systematic quality assurance. They could promote the dialogue within the school and between the school and the environment. The overall idea is that, more than in the past, parents, pupils and students get the opportunity to affect the quality of education.

Publication of School Results in Daily Newspapers

On the basis of the Act “Publicity of Administration” the government can be forced to make available data on school results. Until now the newspaper “Trouw” published an overview of Education Inspectorate data on exam results and transfer figures from secondary education three times (October 1997; June 1998; September 1999).
Another newspaper published an overview of the results of the Integral Supervision of Schools at four hundred primary schools during the first six months of 1998. In ‘Integral school supervision’, schools are examined and evaluated by the Educational Inspectorate on characteristics that are relevant to school quality (both the school’s results as well as quality characteristics relevant to the teaching and learning processes (Sikkes, 1998; Ministry of Education, Culture and Science, 1999).

The Quality Card
In pursuance of the first “Trouw” publication the State Secretary of Education, Science and Culture, Ms. Netelenbos, requested the Educational Inspectorate to develop a so-called ‘Quality card’. This card should contain more quantitative information than “Trouw” so far had published. In addition the specific situation of the school should be considered.

The first publication of the “Quality Cards” was in June 1998. This Quality Card reported among others data on examination results per cluster of subjects per school compared with the profile of comparable schools (Hendriks, 2000; Ministry of Education, Culture and Science, 1999).

The second Quality Card was published in September 1999. On this card also data were presented on the success rate of the first three years of secondary education, with which success rate includes both repeaters and the progress of pupils in relation with the education level recommended when leaving the primary school. And, instead of comparing the results with results of comparable schools as was the case in 1998, at the 1999 Quality Card the school is compared with schools with a comparable pupil population.

The Quality Cards are primarily meant for parents of pupils in the last year of primary education and can be used when choosing a school for secondary education. But the cards could also be interesting for the current pupils of the school for secondary education, their parents, teachers, management and administration. They become informed about the results of their own school in comparison with schools with a comparable pupil population.

The Educational Inspectorate
The changed role of the Dutch Educational Inspectorate has already been referred to in the above. These changes are best illustrated by sketching the developments in primary education. Similar developments in secondary education are being implemented.

In primary education, during the past few years, the supervision system already has undergone important changes (Ministry of Education, Culture and Science, 1999, p. 40). The developments have led to an adapted approach of
supervision, consisting of regular supervision (first order supervision) and integral supervision (second order supervision). From 1999 onwards the Inspectorate works according to this system.

In a regular supervision the Education Inspectorate studies the school’s results and a number of important quality characteristics relevant to the teaching-learning process. In addition, the Inspectorate evaluates the school plan and the school prospectus. To get the information next to the analysis of documents also school visits take place. In general, a school-visit within regular supervision lasts for one day. After the visit the main findings are presented in a report.

If the regular supervision reveals that problems (threaten to) exist, integral supervision will be carried out. If, on the basis of this integral supervision improvements are necessary, then the school will be invited to draw up and implement a plan of action. Depending on the progress, an intensified tailored supervision will be carried out.

A set of explicit standards has been developed for the integral supervision of schools.

Examination and Assessment
Primary education has no final examination in the Netherlands. In the categorical system of secondary schools there is a final examination for each of the four main types of categorical schools. There is a national assessment of all main subject matter areas in primary education, but this is on the basis of a sampling frame that does not allow school-based interpretation of the data. However, about 80% of primary schools uses a partly criterion based assessment test in the final grade which is primarily meant to facilitate the choice of one of the secondary school types. Increasingly this test is also being used to assess the performance of the schools. Some of the major cities have made the administering of the test compulsory for this purpose.

Recently the Education Council, which is the most important advisory body of the Minister of Education and Parliament, has proposed, to use a national compulsory test of the realization of the primary school standards.

Primary schools in the Netherlands are quite used to standardized achievement test, as about 70% of the schools use pupil monitoring systems for purposes of diagnosing the progress of individual pupils. The most frequently used pupil monitoring system, developed by the Institute for Educational Measurement (CITO) is discussed in more detail in a subsequent section of this chapter.

Comparison with other European Countries
When the Dutch context for school self-evaluation is compared to that in other European countries, there is somewhat more central initiative than in countries
like Spain, Italy and other southern European countries. The situation is comparable to that in Belgium, and Scandinavian countries like Sweden and Finland. Phenomena like systematic school evaluation and assessment are relatively new in the context of most of the German Lander, whereas school self-evaluation in France is mostly school oriented feedback on the basis of national monitoring efforts.

Countries where the policy context is more encouraging or even compelling as compared to the Dutch situation are England and Scotland.

In England, gearing school self-evaluation to school developing planning within the context of curricular reform appears to be an interesting combination.

Types of School Self-Evaluation

School self-evaluation can be defined as the type of evaluation where the professionals that carry out the program or core-service of the organization (i.e. teachers and head teachers) initiate the evaluation and take the responsibility for the evaluation and the evaluation results of their own organization (i.e. their classes, the school). “Taking responsibility” may imply that the professionals carry out the school self-evaluation themselves or make use of external advisors.

For school self-evaluation more and more tools and instruments have become available. They can be described and categorized in a number of different ways, among others by the product and process-orientation, the domains and areas of evaluation, the actors involved, the evaluation methods used, the providers, etc. (Scheerens, 1999).

Product and Process Evaluation

Product evaluation is evaluation in which the effects of the ‘products’ of the primary process of an organization are the main targets for evaluation (Scheerens, Hendriks & Bosker, forthcoming). With process evaluation, criteria are drawn up with regard to the transformation processes or support conditions within organizations. This is a questionable form of evaluation leading to reasoning along the lines of ‘operation successful, patient died’. However process evaluation is important as an aid for product evaluation – chiefly in order to gain more insight into the background of certain unexpected or disappointing results.

Domains and Areas for Evaluation

In the European pilot project “Evaluating Quality in School Education” (MacBeath et al., 1998) four main evaluation domains are distinguished, namely
evaluation of outcomes, processes at the classroom level, processes at the school level and the environment. Each domain is subdivided into three areas. In the project 12 areas for evaluation are distinguished.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td>• Academic achievement</td>
</tr>
<tr>
<td></td>
<td>• Personal and social development</td>
</tr>
<tr>
<td></td>
<td>• Pupils’ destinations</td>
</tr>
<tr>
<td>Processes at classroom level</td>
<td>• Time for learning</td>
</tr>
<tr>
<td></td>
<td>• Quality of learning and teaching</td>
</tr>
<tr>
<td></td>
<td>• Support for learning difficulties</td>
</tr>
<tr>
<td>Processes at school level</td>
<td>• The school as a learning place</td>
</tr>
<tr>
<td></td>
<td>• The school as a social place</td>
</tr>
<tr>
<td></td>
<td>• The school as a professional place</td>
</tr>
<tr>
<td>Environment</td>
<td>• School and home</td>
</tr>
<tr>
<td></td>
<td>• School and community</td>
</tr>
<tr>
<td></td>
<td>• School and work</td>
</tr>
</tbody>
</table>

*The Actors Involved in the Evaluation*

Scheerens, in Tiana et al. (1998) distinguishes four main categories of actors in all types of evaluation, including school self-evaluation, namely

- the contractors, funders and initiators of the evaluation;
- the (professional) staff that carry out the evaluation;
- the persons in the object-situation which provide data;
- the clients or users or audiences of the evaluation results.

When looking more in-depth to the providers of data, several groups of respondents can be distinguished (i.e. pupils, teachers, head teachers, other staff, parents, schools for secondary and further education, supply schools, etc.).

*The Methods Used*

A distinction can be made between quantitative and qualitative approaches. With quantitative procedures, the approaches for data collection must conform to requirements concerning validity and reliability (Scheerens, Hendriks & Bosker, forthcoming). Qualitative approaches have the following characteristics:

- use of "open" research formats, such as "open" interview questions, and "free" observation;
- a strong dependence on the views of persons that are part of the "evaluandum" (the evaluation object);
School Self-Evaluation in the Netherlands

- narrative, and sometimes so-called “thick description” of the object situation rather than quantitative output (tables, graphs);
- smaller aspirations towards generalizibility of findings because of the fact that fewer units or codes are studied “in depth” (Scheerens in Tiana et al., 1999, p. 62).

Qualitative and quantitative approaches each have their strong and weak points, and sometimes, a combination is the best solution.

Providers
In the Netherlands, there is a broad supply in approaches and tools for school self-evaluation. There is also a broad and very differentiated supply in providers of tools and services with regard to school self-evaluation. In a study under the authority of the Dutch Ministry of Education, Culture and Sciences (Hendriks, 2000) 44 providers in the field of quality assurance and school self-evaluation in secondary education and their main products (instruments) and services were listed. The providers can be distinguished into the following categories:

- the education support services;
- the so called “Process Management” for secondary education (PMVO);
- the national educational advisory centres (LPC): the Educational Advisory Centre (APS) for non-denominational schools, the Protestant Educational Advisory Centre (CPS) and the Catholic Educational Advisory Centre (KPC);
- the Institute for Educational Measurement (Cito);
- the Institute for Curriculum Development (SLO);
- the European Platform for Dutch Education
- the national union of school students (LAKS);
- the national parents’ organizations;
- the association of school leaders in secondary education (VVO)
- the school board unions;
- the education unions;
- the foundation for replacement and business health care for the education sector (in Dutch: “Stichting Vervangingsfonds en Bedrijfsgezondheidszorg voor het onderwijs en Stichting Participatiefonds”);
- the in-service training departments of the teacher training institutes;
- the semi-private institutes affiliated to universities and colleges of higher education;
- private firms.
The report further offers in-depth description of 16 instruments for school self-evaluation.

A Closer Look at Specific Types of Instruments and Providers
Since it is impossible to give a description of the whole range of data-collection methods and types of instruments for school self-evaluation and the whole range of providers, only three methods of data-collection will be discussed briefly, namely:

• the Dutch Quality Model of the Institute for Dutch Quality;
• “Review Secondary Education” of the three national educational advisory centres (LPC) in collaboration with the association of school leaders in secondary education (VVO), and;
• the Checklist to develop “a Golden School” with the “Livable school questionnaire” as its major instrument of the Dutch national union of school students (LAKS).

The Dutch Quality Model (MNK Model)

The Dutch Quality Model (Model voor Nederlandse Kwaliteit) (MNK Model) is the Dutch version of the EFQM Excellence Model of the European Foundation for Quality Management (EFQM). The Institute for Dutch Quality (Instituut voor Nederlandse Kwaliteit) adapted the EFQM model for the Dutch situation, mainly by elaborating the model with five developmental stages.

The Dutch Quality Model is a non-prescriptive framework aimed at stimulating organizations both in the profit and non-profit sector to “continuously” improve the quality of their products and services. Four different versions of the model are available (respectively for companies, the government, and the fields of education and health).

So far in education, the MNK model is mostly used in Higher Vocational Education, in Secondary Vocational and Adult Education. Use in Secondary Education is also getting off the ground, nowadays.

A sketch of the structure of the MNK model is given in Fig. 1.

The Dutch Quality Model consists of nine criteria against which an organization can assess its progress towards excellence. All the criteria are interconnected and of the same value. The model is divided into two parts, namely organization and results. The following criteria are distinguished:

• in the field of organization: Leadership, Policy & Strategy, Human Resources Management, Resources Management, Management of Processes, and;
School Self-Evaluation in the Netherlands

Leadership

Human Resources Management

Policy & Strategy

Management of processes

Customer satisfaction

Staff Member Satisfaction

Impact on Society

Resources Management

Business Results

Organization

Results

Feedback

Fig. 1. The Dutch Quality Model (Stevens, 1998).

- in the field of results: Staff Member Satisfaction, Customer Satisfaction, Impact on Society, Business Results.

The two parts are closely connected. Each organization will first of all be tending to examine the results. The level of the results is seen as a first indicator on the proper functioning of the organization.

Each criterion is supported by a number of subcriteria. The sub-criteria pose a number of questions that should be considered in the course of an assessment.

Five developmental stages indicate a path of growth in which each stage implies the previous developmental stages:

(1) Activity-oriented
A good performance of the activities are the central issue, much emphasis is placed upon the rectification of problems once they have occurred;
(2) **Process-oriented**

The primary processes and the management of the process form the central issue; improvements are implemented in problem areas on the basis of measurement and knowledge of the processes;

(3) **System-oriented**

The total organization is controlled, including the supporting departments; the management of all processes is governed by internal and external customer orientation, the goal is to pre-empt problems and complaints;

(4) **Chain-oriented**

Maximum use is made of the knowledge and capacities of the organization in relation to suppliers, customers and other interested parties in order to satisfy the customer/target group, in conjunction with these, the most effective distribution of tasks is worked out and win-win situations are pursued and realized;

(5) **Total Quality Management**

The vision and policy of the organization are formulated with the intent of declaring responsibility to society. The total quality management is anchored, both internally and externally (Degenaar & Van Kemenade, 1998, pp. 4–6).

With regards to each criterion and at each stage the principle of continuous improvement is promoted.

The Dutch Quality Model can be used to make a self-diagnosis and to formulate points for improvement. In addition an external certified INK auditor could be asked to verify the self-assessment.

Next, support in all the steps of the assessment can be enlisted in the form of a growing number of consultancies, the so-called ‘associates in knowledge’ of the Institute for Dutch Quality. Only these, by INK certified auditors are authorized to use the materials of the Institute for Dutch Quality and to take care of the requested support.

*Secondary School Audits*

Audit grounded in self-review is a method for quality development of the school. It consists of a self-assessment by the educational organization and a visit/audit of the organization by an external committee of independent experts. It is also possible that a group of schools review each other under the leadership of a review committee of independent experts (fraternal review).

In Dutch education, the self-review and audit approach is mainly used in Higher Vocational Education. In secondary education in 1996/1997, 18 schools
in the province of Limburg, all belonging to the same administration, reviewed each other (Steneker, 1997). At present Secondary School Audits are offered as a method for quality improvement by the three national educational advisory centres (LPC) in collaboration with the association of school leaders in secondary education (VVO).

Usually, teams of three schools review each other under the leadership of a review committee of independent experts.

This approach encompasses five themes which are all connected with the education process of the school, the process of guidance and counseling, and human resources management.

Therefore the educational advisory centres (LPC) and the association of school leaders in secondary education (VVO), have set up a list of possible themes. In consultation with the external review committee each school determines the five themes to be reviewed. Preferably the three schools choose the same themes.

The course of review takes seven months and consists of the following steps:

(1) **Set up the review committee**

The review committee is responsible for the content and procedures. The review committee consists of 6 persons: the school leaders of the three schools and three by LPC and VVO invited external members, two of them even from outside the education sector. The third external expert is the committee chairman (from the LPC or VVO);

(2) **Self-assessment**

With the help of a checklist the school makes the assessment on the themes chosen and prepares a written report containing information about the schools' intentions, actual functioning and judgments;

### Table 1. Overview of Possible Themes.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education process</td>
<td>Implementation of the innovations in a secondary education (i.e. the introduction of the so-called “basic secondary education” and the changes in the top years of secondary education, active and independent learning, etc).</td>
</tr>
<tr>
<td>Coaching process</td>
<td>Coaching of pupils at entry, moving up and leaving the school, coaching of pupils during lessons, pupil monitoring system, an orderly and safe school climate, etc.</td>
</tr>
<tr>
<td>Human resources</td>
<td>Professional standards, professional development, reduction of cancelled lessons, assessment, teambuilding, cooperation, etc.</td>
</tr>
</tbody>
</table>
Audit
Based on the findings in the written report, the review committee visits the schools and conducts dialogues with the different groups concerned: i.e. students, teachers, non-teaching staff, school management, parents, etc.

Review report
In view of their results and analyses the committee formulates a review report, including also recommendations for improvement;

Starting improvement activities
The school prepares an improvement plan.

Checklist to Grow to a "Golden School";
the Livable School Questionnaire

The Dutch national union of school students (LAKS) is an organization of students (and schools) in secondary education. LAKS looks after the interests of students by providing information to students, schools and government and representation in several discussion groups and organizations, both at the regional and country level. LAKS is the only organization of students recognized as a forum for students by the press, the authorities and the educational organizations.

Checklist to Grow to a Golden School
In 1995 LAKS presented the ‘Checklist to grow to a golden school’. The checklist was aimed to raise the awareness of students about the performance of their school and the possibilities to improve the quality of their schools (OBESSU, 1998). According to LAKS a school has to be ‘a place of comfort for young people, in which they can feel safe, feel respected, are being stimulated to learn and acquire knowledge and skills, and where they can participate in the structuring of their day to day life’.

The ‘Checklist to grow to a Golden School’ is aimed to provide students with a step-by-step handbook to evaluate and improve different aspects of their school.

Golden School Survey
In 1996/1997 LAKS conducted the ‘Golden School Survey’ survey among 20,000 third grade students in 100 schools for secondary education. In total 13,500 students of 86 schools completed the questionnaire and sent it back to LAKS.
The questionnaire contained 24 questions about, among others, the social climate, the interaction of students and teachers, the quality of teaching and learning and the quality of the facilities at school and the building and environment.

The results were processed with the help of the educational services and research department of the University of Nijmegen (IOWO) and the Process Management for Secondary Education (PMVO) and presented in a final report.

*Livable School Questionnaire*

In 1998 LAKS developed the ‘Livable school questionnaire’. The ‘Livable school questionnaire’ is a list with requirements to which an ideal livable school should meet. With the ‘Livable school questionnaire’ the school can be judged first and then improved step-by-step. The questionnaire contains 50 questions divided among the following topics: school culture, lessons, coaching, interaction, active involvement of pupils, provision of information, discipline and safety, school building and environment, organization and students’ participation.

The questionnaire is available on the Internet (http://www.laks.nl). It can be completed electronically, but it can also be downloaded from the Internet at no costs and used in paper and pencil format. The questionnaire can be completed within half an hour.

The questionnaire starts with a short explanation and an overview of the 10 topics included. More in-depth information can be obtained by touching the specific topic. Then the 50 questions can be answered. After completing the questionnaire the results (i.e. the livability of the school according to the respondent) are presented automatically, both in numbers (percentages) and graphically.

The topics with a bad score should be elaborated further and improved. By ‘touching’ these topics and/or the specific questions per topic the respondent automatically receives information and suggestions how to improve the livability of the school with regard to this topic.

**A BRIEF REFLECTION ON THE DUTCH SETTING REGARDING SCHOOL SELF-EVALUATION**

The general political climate in the Netherlands can be seen as “mildly supportive” of the further development of school self-evaluation. There are generally favorable and stimulating conditions that are even enforced by formal requirements. Yet, there are not yet “hard edges” in external incentives to
schools to present themselves to constituencies, as may be the case when school self-evaluation is seen as a counterbalance against external assessments, although the publishing of examination results in newspapers may start to create this effect.

The providers, methods and tools that were illustrated are a reflection of the Dutch education support structure on the one hand and a beginning penetration of the “quality care business” in the education sector on the other. It is important to note that both orientations have more of a “social engineering” background than a grounding in empirical analytic design and evaluation methodology. In terms of well-known standards to assess evaluations, these approaches are stronger in criteria like utilization focus than they would score on accuracy standards. Colloquially speaking these approaches could be called “soft”. Evaluation studies on school self-evaluation approaches, funded by the Ministry of Education, noted a general lack of attention for the reliability and validity of instruments for school self-evaluation (Cremers van Wees et al., 1996a, b; Hendriks, 1999).

The initiation of the instrument development project that is the focus of the second part of this chapter, the ZEBO-project, is to be seen as a direct consequence of these evaluation studies. At the same time the more empirically analytically inclined branch of the Dutch education support structure, particularly the Institute for Educational Measurement (CITO) and the Foundation for Educational Research were also interested in making their mark on the scene of school self-evaluation. The result was a collaborative research project between CITO, OCTO, the research institute attached to the Faculty of Education of the University of Twente and the SLO, the Foundation for Curriculum Development, aimed at what was later termed an “integrated” instrumentation of school self-evaluation for primary schools. This project is described in detail in the remaining part of this chapter.

THE ZEBO PROJECT

As stated above, the project was initiated and funded by SVO, the Dutch Foundation for Educational Research in co-operation with CITO (Central Institute for Test Development) and SLO (Foundation for Curriculum Development). The qualification “integrated” refers to the fact that different types and approaches to school self-evaluation and monitoring educational quality are being combined. These various approaches also have distinct theoretical and disciplinary backgrounds, which in a way are also synthesized
in this project. Briefly stated, the joint instrument will consist of: (a) a pupil monitoring part that depends heavily on psychometric theory and central issues of adaptive instruction; (b) an assessment of the educational content covered (also often indicated with the term “opportunity to learn”) with central concepts of curriculum planning and curriculum evaluation at its background; and (c) the measurement of school process indicators, with school effectiveness and school improvement modeling as its conceptual background.

Of these three types pupil monitoring systems in particular deserve some further explanation. The focus of attention in this chapter is self-evaluation at the school level. Pupil monitoring systems operate at the micro level (class level) of educational systems. In the ensuing sections it will be shown how this class of techniques can also be used for self-evaluation at the school level.

Basically pupil-monitoring systems are sets of educational achievement tests that are used for purposes of formative didactic evaluation. An important function is to identify those pupils who fall behind and where they experience difficulties.

Pupil monitoring systems have one asset which, in our opinion, is essential for all efforts to make school functioning more effective: the centrality of output data at the level of the individual pupils measured by means of achievement tests. If approaches to school self-evaluation neglect these type of data there is a risk that the information basis they supply for educational or administrative decision-making is faulty.

The other two components will be further described in subsequent sections.

In the most general sense the instrument (or set of instruments) being developed should enable schools to assess their own quality. In its turn this quality assessment can serve three major purposes:

(a) to provide information for quality maintenance and school improvement, or to use a fashionable term, to provide a basis for “organizational learning”,
(b) to provide a basis for informing relevant audiences (for instance parents) about school functioning,
(c) to provide a starting point for further analysis and diagnosis on specific points for further aspects of school functioning.

General Characteristics

Educational quality has many facets. In the design of instruments aimed at quality assessment choices and the setting of priorities cannot be avoided. The
first choice is to develop instruments for school self-evaluation that comprise information on inputs, processes and outcomes. Inputs are both financial and material resources and the "human resources" that pupils provide when entering a school. In other words pupil background characteristics like socio-economic status and initial achievement should be measured. A second choice is concerned with the interrelationships between inputs, processes and outcomes. Here the orientation chosen is to be outcome directed. This means that measurements of educational outcomes are the backbone of the instrument. Input data concerning pupil background characteristics are used to construct "value-added" outcome measures and process indicators including those concerning the content that is covered are used to provide ideas and "hunches" as to what accounts for certain patterns in value-added outcomes.

In developing the school self-evaluation instrument an important further decision -given the centrality of outcome measures- concerns the question as to which outcome measures should receive priority. Priority in the sense that one cannot do everything at once, and, "for practical reasons only", a start should be made with just one or two outcome areas. In our case the decision was taken to focus first on basic school subjects like arithmetic, language and reading. Without denying the fact that other educational outcomes are also important, this choice can be justified by pointing out the important place these subject-matter areas have in the primary school curriculum. Besides, there was a more practical consideration: in these areas outcome measures validated and normed achievement tests are already available in the Netherlands.

The next choice to be made concerned the type of input and process measures to be included in the instruments. Given the need to adjust "raw" outcome scores for prior attainment and other pupil background characteristics in order to obtain information on the value-added by schooling, such variables need to have a place in the design of the overall instrument. Ideally one would like to have longitudinal achievement measures in the core subject-matter areas throughout the duration of a school career, in order to monitor the progress that pupils make in these areas closely. In this aspect the CITO-pupil monitoring system (cf. Moelands & Ouborg, 1995, Gillijns & Moelands, 1992) that is available in the Netherlands provides excellent opportunities. In this system, from Grade three onwards there are bi-annual achievement tests which have been "vertically equated", which means that scores at one point in time can be compared to scores at a further point in time. Apart from these "general" tests that are designed to have a signaling function with respect to pupil progress there are also diagnostic tests to be used as a follow-up for individual pupils if there is a need to do so.
The rationale adopted in the choice of process variables was to select those variables that, in educational effectiveness research, have been shown to be associated with relatively high "value-added" achievement.

Considering the types of variables chosen for the overall school self-evaluation instrument and the general methods of measuring them, the overall design of the instrument, indicated in Table 2 shows which components (variables and methods) are meant to be integrated into one set of instruments.

Although "integration" is one of the guidelines of this whole project it should nevertheless be further qualified, because the connections that can be made between certain elements of the overall instrument will be closer than of other combinations. For instance, achievement and prior achievement data will be combined into value-added indicators or progress-scores. In principal it is also possible to provide measures of association between content-covered and achievement. In general, but dependent on the interpretative frameworks and procedures of analysis to be discussed in a later section, there will be a looser association between achievement indicators and process-indicators, particularly when these are measured at the school level. Maybe the instrument to be developed should also provide options for schools to use it in a modular form, so that each of the parts (see the right-hand column of Table 2) can be used on its own.

Another optional feature of the instrument could be to make a distinction between basic monitoring and signalling functions on the one hand and additional instruments for further analysis and diagnosis. For the school monitoring part this distinction already exists. But this idea might also be generalised to the area of process-indicators, for example by providing more detailed instruments to assess school management.

Table 2. Components to be Integrated in School Self-Evaluation.

<table>
<thead>
<tr>
<th>Type of Variables</th>
<th>Sub-Categories within Types</th>
<th>School Self-Evaluation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>input indicators</td>
<td>- initial achievement</td>
<td>School management</td>
</tr>
<tr>
<td></td>
<td>- other background</td>
<td>information Systems</td>
</tr>
<tr>
<td></td>
<td>characteristics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- financial and material</td>
<td>Curriculum evaluation</td>
</tr>
<tr>
<td></td>
<td>inputs</td>
<td>School diagnosis</td>
</tr>
<tr>
<td>process indicators</td>
<td>- content covered</td>
<td>instruments</td>
</tr>
<tr>
<td></td>
<td>- effectiveness enhancing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>conditions</td>
<td></td>
</tr>
<tr>
<td>outcome measures</td>
<td>- achievement adjusted for</td>
<td>pupil monitoring systems</td>
</tr>
<tr>
<td></td>
<td>initial achievement</td>
<td></td>
</tr>
</tbody>
</table>
Variables Included in the Over-All Instrument

Although the final decision has not been taken yet, the most likely selection of input, process and outcome variables to be included is shown in Table 3.

Table 3. The Most Likely Selection of Variables to be Included in an Integrated Instrument for School Self-Evaluation.

<table>
<thead>
<tr>
<th>Outcome variables</th>
<th>Arithmetic, language/reading achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior attainment variables</td>
<td>Prior attainment in these subjects</td>
</tr>
<tr>
<td>Other student background variables</td>
<td>Socio-economic status</td>
</tr>
<tr>
<td></td>
<td>School career data (e.g. repeating grades)</td>
</tr>
<tr>
<td></td>
<td>Pupils entering the school system at a later age than legally prescribed (e.g. pupils from foreign countries)</td>
</tr>
<tr>
<td>Content covered</td>
<td>Methods (school books) used in core subjects</td>
</tr>
<tr>
<td></td>
<td>Components of subject matter instructed</td>
</tr>
<tr>
<td></td>
<td>Opportunity to learn (e.g. test-items covered at 2 points in time during school year)</td>
</tr>
<tr>
<td>Achievement orientation, high expectations</td>
<td>General achievement orientation</td>
</tr>
<tr>
<td></td>
<td>Consensus on educational objectives</td>
</tr>
<tr>
<td></td>
<td>Teacher expectations</td>
</tr>
<tr>
<td>Educational Leadership</td>
<td>Frequency of educational leadership tasks</td>
</tr>
<tr>
<td></td>
<td>Main performers of educational leadership tasks</td>
</tr>
<tr>
<td></td>
<td>Staff professionalism</td>
</tr>
<tr>
<td>Consensus and cohesion among staff</td>
<td>Frequency and content of formal staff meetings with the school management</td>
</tr>
<tr>
<td></td>
<td>Frequency and content of informal meetings among teachers</td>
</tr>
<tr>
<td>School and classroom climate</td>
<td>Relationships between pupils</td>
</tr>
<tr>
<td></td>
<td>Relationships between teacher and pupils</td>
</tr>
<tr>
<td></td>
<td>Relationships between staff</td>
</tr>
<tr>
<td></td>
<td>Relationship: the role of the school head</td>
</tr>
<tr>
<td></td>
<td>Working conditions and task load</td>
</tr>
<tr>
<td></td>
<td>Order and work attitude</td>
</tr>
<tr>
<td>Instruction</td>
<td>Structured instruction</td>
</tr>
<tr>
<td></td>
<td>Monitoring of pupils’ progress</td>
</tr>
<tr>
<td>Pupil care</td>
<td>Pupil care at the school level: measures that enable the realisation of inclusive education</td>
</tr>
<tr>
<td></td>
<td>Pupil care at the classroom level: adaptive instruction</td>
</tr>
</tbody>
</table>
Interpretative Frameworks and Analyses

When it comes to the analysis and interpretation of the data to be obtained by our integrated school self-evaluation instrument there are two major analytic requirements.

(1) to provide "net" or "value-added" measures of school-performance;
(2) to attribute variation in these value-added outcome-measures to process variables.

In the situation where the same data are available from a sufficiently large group of schools these two problems of analysis can be of resolved by using appropriate multi-level modelling techniques (cf. Bosker & Scheerens, 1995).

In fact, comparative information on a large number of schools is essential to provide an interpretative basis for the performance of each individual school. This does not mean that school self-evaluation would only be possible in a "permanent" comparison of schools, but merely that it would be desirable to establish norms on the basis of a sufficiently large reference group of schools. Even if one were to think of using panels of experts to establish "absolute" performance norms (e.g. a particular cutting score on a scale of adjusted outcome measures), one would like to have information on the actual performance of schools, before setting such norms. So, in theory schools could distinguish three types of interpretative frameworks to evaluate their performance:

(a) by referring to nationally established norms;
(b) by concurrent comparison with other schools;
(c) by interpreting a school's performance by means of "introspection" (no external referents).

In the latter case a school could keep track of its performance over time and see if changes could be seen as improvement or as falling back. Moreover, a school could set specific attainment targets in terms of specific increments on "net" value-added performance. Although this would be quite a sensible procedure, it is expected that schools would be also interested in some kind of external referent when applying such norms, which underlines the need for comparative information.

When it comes to the second analytic problem, that of attributing variation in performance to process or input variables, there are additional difficulties.

There are in fact two ways for estimating "scores" on process-measures. In the first place one could think of acceptable levels for each process indicator. For instance, one could say that an acceptable level of instructional leadership would be reached when a head teacher would spend 20% of his or her time...
on instructional as opposed to other managerial tasks. The relevance of such statements, however, would depend on the importance of such a process-indicator for enhancing school performance. Ideally, to meet this latter requirement, one would be interested in the precise association of the score on a process indicator and the level of performance. To the degree that such coefficients would exist for major process-indicators, one could say that the educational production function of schooling would be known. It is doubtful whether the establishment of something like an educational production function is at all realistic (cf. Monk, 1989). In any case, although we know quite a lot about "what usually works" in primary education, the current knowledge base lacks the precision and consistency to come anywhere near a complete instrumental explanation of educational performance. This implies that on-site knowledge is indispensable in interpreting process-output connections when carrying out school self-evaluation.

Although external referents on acceptable levels of process-behaviour would be most helpful, the creativity and inside-knowledge of school-teams would be required to contemplate particular performance patterns.

School internal comparative information on process-characteristics is to be seen as a basic prerequisite for making any tentative attributions with respect to outcomes. So, a school would benefit most from comparing process "scores" or descriptions over time or (when there are parallel-classes), between classes and teachers.

**Data Requirements**

The integrated school self-evaluation instrument contains different types of variables and different types of methodology to measure them. In order to be able to carry out the analyses that were globally referred to in the preceding section specific requirements concerning the data and the design to collect data should be used. In addition, there are differences in the state of the art of instrumentation and measurement among the various types of variables. A closer look will be taken at outcome data and the conduct of value-added outcome indicators, instrumentation of opportunity to learn and the measurement of process indicators at school and classroom level.

**Output Measures**

Psychometric theory applied to educational achievement measurement is a highly developed discipline. Item response modelling in particular has improved the unambiguity concerning results of achievement tests. For our purposes this gain in unambiguity is particularly relevant with respect to the measurement of
progress. When achievement tests administered at various points in time confirm to the assumptions of item response-models, it becomes possible to draw reliable conclusions on gain or progress in the achievement of individual pupils and groups of pupils. For further details on the application of item response theory in pupil monitoring systems the reader is referred to Gillijns and Verhoeven (1991).

The psychometrically strong properties of achievement tests that confirm to item response-models are also advantageous in computing so called value-added indicators (cf. Thomas, 1992)\(^2\). The basic idea of value-added measures of schooling is a comparison of a predicted score based upon past performance with actual performance. Depending on whether the difference (technically referred to as a *residual*) is positive or negative the pupil’s performance is seen as over- or under-achievement. By aggregating residuals over pupils the over- or under-achievement of a school or classroom can be determined. In school effectiveness research value-added outcome indicators are to be used as the dependent variables.

*The Measurement of Content Covered*

The aim of measuring the subject matter content that is covered over a particular period of tuition involves checking the degree to which pupils have actually been exposed to the subject matter that is required in tests or examinations. In internationally comparative assessment studies “opportunity to learn” is used as a control variable in order to be able to make “fair” comparisons of educational achievement between countries. In school effectiveness research, variance in content covered is seen as one of the process characteristics that can explain differences in achievement between schools.

Within the context of an integrated school self-evaluation instrument, the function of periodical measurement of content covered would be to check to what degree tuition in a particular subject has been targeted to the objectives. Less exposure to certain subject-matter elements might explain a falling back in achievement in a particular classroom or in a particular school year.

There are various ways to conceptualize and measure content covered. An important conceptual distinction is between merely measuring subject matter elements on the one hand and adding additional qualifications like time of exposure and correspondence between presentational aspects and the way test questions are asked. Here the first, more limited interpretation will be used.

Next, there are different data sources that could be used in measuring content covered. The teaching methods (schoolbooks) used could be one source. Content analysis of books and other written teaching material and an indication as to “how far” a teacher had progressed through this material in a particular period could then be the basis for a measure of content coverage. Another likely data-source
would be teachers, by asking them in a structured way about the content covered in a particular subject over a certain period. And a third, quite detailed way to measure content covered is to ask pupils to indicate whether or not a particular test-item has been taught. Each of these potential data-sources and data-collection measures has its particular strong and weak points in terms of practical feasibility and methodology. These will be systematically considered in making a particular choice with respect to the way content covered could best be measured within the framework of school self-evaluation. Practical feasibility and efficiency in procedure are particularly important criteria in making this choice.

Next to this, from the perspective of curriculum theory a fundamental issue surrounding the measurement of content covered is the degree to which consensus can be reached about the distinction and sequence of core content-elements that are thought to be necessary to attain particular educational objectives. However, important differences are likely to exist between subject-matter domains. In the case of applying content covered measures in a school self-evaluation instrument such core content-elements should also be acceptable and credible for the schools.

In a study carried out by the Institute for Educational Measurement (Cito) a questionnaire has been drawn up and presented to 170 teachers of grade 3. All these teachers make use of the Cito-pupil monitoring system. In this study, the development of an instrument to measure content covered is limited to arithmetic. In the study the following three research questions are at the centre:

• Is it possible to develop a reliable and valid instrument to measure content covered?
• If yes, what should be the design of such an instrument?
• To what extent does curriculum overlap (content covered) exist?

This last question is important since the tests from the Cito-pupil monitoring system are independent of the methods (schoolbooks) used in the schools.

While there are various education stakeholders that can be consulted about the measurement of content covered (i.e. pupils, teachers), for this research teachers were chosen as the target group.

Because different data sources could be used in measuring content covered, the questionnaire sent to the teachers comprised three parts:

• one part was related to the method (school books) used in arithmetic;
• the following part was related to the components of subject matter instructed and;
• the last part was based on opportunity to learn (test-items covered).

At this moment the data are being analyzed thoroughly. It is expected that analyses will show what working method will be best suited for the measurement
of content covered, in which practical requirements such as the investment time and such psychometric requirements as dimension, validity and reliability are taken into account. It will also be explored whether it is possible, with the help of an item response theory, to develop a content covered scale that gives a quantifiable pronouncement on the extent to which tasks should be explained to pupils. Finally, on the basis of available pupil responses to the tasks and methods used by the teachers, the possibility of method bias will be investigated.

The way in which the content covered instrument will be included in the overall set of instruments will depend on the results of the study. It is thought that the content covered instrument can be included as an explanatory variable for a multilevel model, still to be developed. In the Cito-pupil monitoring system raw scores with conversion tables are translated into an item response theory constructed proficiency scale. A second possibility is that adjusted conversion tables of raw scores are developed into proficiency scores. It is then no longer necessary to include content covered as a declared variable in the proposed multilevel model.

*The Measurement of Process-Indicators*

From the literature on school effectiveness, school improvement and on existing methods for school-diagnosis a set of most relevant process-variables can be obtained.

At the same time there are only a few more standardised instruments available to measure these activities. There appears to be a strong tendency for researchers to design their own instruments for each and every new research study.

An initial task has been the making up of an inventory of available instruments. In this way an attempt has also been made to capture the operational core that are usually mentioned in the reviews of school effectiveness research, in order to get insight into the actual conceptual contents of complex constructs like “educational leadership” and “high expectations for students’ achievement.” Nine (inter) national school effectiveness studies and five Dutch school self-evaluation instruments were used as the basis for this inventory. Thirteen general factors often mentioned in school effectiveness research were analysed and the elements found in the operational definitions and instruments concerning these factors were summarised for each factor. (The reader is referred to Scheerens and Bosker (1997), for further details). The inventory has led to a broad range of components within factors. At the operational level it has been turned out that there is little agreement on the substance of key factors that are supposed to determine school effectiveness. Then the factors and components found were presented to primary school leaders and teachers. Partly on the basis of their opinions, but also among others on the basis of the conclusions of the
Committee for the Evaluation of Primary education (1994) a first selection of process-variables was made. 

Another problem concerns the reactivity (e.g. the fact that respondents are in the position to "colour" responses) of the most frequently used methods to measure process indicators. On the other hand the reality of process characteristics that are thought to be relevant with respect to quality is not just factual but also "opinionated." Expectations, for instance, may be highly subjective in origin, but may nevertheless create factual conditions for pupils. Nevertheless all possible ways to avoid biases in measuring process indicators will be examined. One possibility could be to avoid self-reporting as much as possible and to question pupils on teachers, teachers on head teachers and vice versa.

To avoid cumbersome procedures this part of the instrument should take the shape of a limited set of relatively early measurable core-indicators for standard use and a handbook containing more detailed instruments and procedures for more in depth organisational diagnosis.

*The Instrumentation 'School and Class Characteristics'*

Based on concepts from the research into school- and instructional effectiveness the University of Twente has developed instruments to measure the process characteristics achievement orientation, educational leadership, staff development, consensus and cooperation, school/class/working climate, structured instruction and adaptive instruction. The idea of the instrumentation is self-reporting and judgment by "consumers". So, double measures are taken: at the school level principals provide information using self-reports and are also judged by teachers, and at the classroom level teachers principals provide information using self-reports and are also judged by their pupils. Three instruments are available: one for the school leader, one for the teachers and one for the pupils.

*Examples of items of the pupil questionnaire*

*Mostly I have enough time to finish my work*

- True
- Bit true
- Not true

*I think many children in my class like me*

- True
- Bit true
- Not true
In a first phase, the instruments were tested in 43 primary schools. Forty-four school leaders, 396 teachers and 2744 pupils in grade 4, 5 and 6 completed the questionnaires. Afterwards each school received a feedback report with information about the school as a whole and information about the individual classes. Also comparisons with the results of the total group of schools involved have been reported.

Example part of the feedback about the school as a whole

Example of feedback school report

From the school overview it appears that teachers from the example school in comparison with teachers from other schools are above-average positive about the educational leadership at their school. The workload compared with other schools is below average and all teachers agree with this (see width of white bar). The scores for staff development are somewhat lower than those from other schools. From the school overview we can see that the average score on co-operation at the example school is very low and that the teachers have different opinions about this subject (see width of white bar).

The aim of the test was to determine the reliability, validity and utility of the instruments. The psychometric aspects involved the estimation of the internal consistency of the scales and aggregates (using multi-level statistical models; Goldstein, 1995), the convergent and discriminant validity and the predictive validity of the instruments.

The investigation of the utility focussed on the completeness and relevance of the topics, the investment of time and easiness of the questionnaires, the feedback and usefulness of the information. For the rest, a lot of attention is paid to anonymity and motivation. The question of utility was answered by in-depth interviews at 8 schools with school-leaders, teachers and pupils in grade 4, 5 and 6. Four schools were selected on the basis of extreme scores of the measured process characteristics. The four other schools were selected because of notable discrepancies in answers by actors from the different levels.
Most instruments appear to be reliable, and what is more, the measurement of aggregates using, for instance, pupil judgement scores on direct instruction behavior of teachers is reliable as well. The correlation of average pupil judgements and accompanying variables measured by teacher self-reports is poor, indicating absence of a relationship between the two methods. This is also true for aggregated teacher scores and principal self-reports, be it that high correlations were found for team cohesion, educational leadership and school leadership. In this case there is a high correspondence between the two methods, indicating concurrent validity.

Explorative analyses to get insight in potential predictive validity of the key concepts, results in low correlations between process measures and pupil progress in arithmetic and language.

Schools are enthusiastic about the instrumentation, especially the pupil questionnaire. The topics are relevant and complete. A supplement that schools would like to have to the system concerns a 'pupil-monitoring-system' for the affective domains. Schools appreciate the form and content of the feedback. The feedback was more specific and clear than they expect. Moreover, the discrepancy information (differences in judgement between pupils and teachers respectively between teachers and principals) appeared to be useful in itself according to the schools.

**CONCLUSION AND APPLICATIONS**

In the initial part of the chapter a sketch of the Dutch policy context with respect to quality care and school self-evaluation in education was given. The particular context of fairly autonomous schools, some recent policy measures and the availability of a range of providers of tools and instruments have led to a broad mixture of approaches to school self evaluation. This context should also be seen as the basic “feeding ground” for the initiative to develop integrated instrumentation for school self-evaluation in primary education, the ZEBO-project.

Next, the theoretical background and methodological aspects of this integrated school self-evaluation instrument have been discussed. The various developmental tasks to actually create such an instrument are still in full progress in the joint project of CITO, SLO and OCTO.

In this final section an outline will be given of the actual uses schools can make of the envisaged integrated instrument. This is done by listing the type
Table 4. Overview of Key-Questions to be Answered by Means of an Integrated School Self-Evaluation Instrument.

A. Are attainment levels in basic subjects improving or declining over time? (within schools, within cohorts “raw score” comparisons)
B. Is the schools’ effectiveness improving or declining over time? (as A, using adjusted, value-added outcomes)
C. Is the schools’ effectiveness up to standard from year to year? (comparative analysis on value-added outcomes)
D. What can be concluded about teacher effectiveness by examining average progress-scores of particular cohort grade-level combinations? (within schools, between cohorts and between grades analysis on progress scores)
E. Is the school differentially effective for certain sub-groups of pupils? (e.g. comparing high and low SES-pupils)
F. How do patterns of differential effectiveness develop over time? (differentiation according to grade-levels and cohorts)
G. Do amounts of content covered in basic subjects increase or decrease over time? (analyses per subject, grade-level and cohorts)
H. How do the school’s content covered levels compare to external referents? (analyses per subject, grade-level and cohorts)
I. Can changes in effectiveness be attributed to variation in content covered? (analyses per subject, grade-level and cohorts)
J. Similar questions as G, H and I for other process indicators on school and classroom functioning

of evaluative conclusions that schools are able to draw when the instrument has been used over a period of say five years and when external reference information is available. A possible list is given in Table 4.

Apart from these basic questions of quality monitoring the use of the integrated instrument may function as a starting point for more in-depth diagnostic investigation requiring specific procedures and additional instruments.

Now that the ZEBO instrumentation study is almost finished the implementation and practical application by the schools is a major focus in completing the project. Several try-outs in pilot schools are envisaged. A follow-up project, which will be fully focused at a complete “computerization” of the instruments and tools is currently being discussed with government officials and municipalities of major cities.

NOTES

1. Since 1997 merged into NWO, the Dutch Organization for Scientific Research.
2. If only because of the fact that the reliability of such measures take away some of the uncertainty surrounding possible components of the residuals “value-added” analyses.
REFERENCES


Department of Evaluation and European Commission, D. G. XXII, Education, Training and Youth, Directorate A Unit 1.


INTRODUCTION

The Austrian school system has traditionally been very centralised. Within the classroom, however, Austrian teachers have been granted freedom of choice in teaching, whatever they choose as relevant within the open framework of the national curriculum and whichever way they find it most appropriate to do it. There has been very little control from outside apart from the inspectors of schools who very rarely “inspect” individual teachers during their careers.

Similarly to other European countries, in Austria schools have been granted more autonomy whereby the school system might become more democratic (i.e. giving more involvement to teachers, students and parents in the decision making processes of the school). In this respect this would give some support to teacher/parent groups which have campaigned for school development from the grass roots since the late 1970s. The expectation is that increased autonomy will help to neutralise the impact of political influences on the educational system, and the Ministry of Education and Cultural Affairs has promoted widespread public discussion on school autonomy. Decision making power over school budgets has been devolved to some schools on an experimental basis
and schools now have some power to take decisions concerning the curriculum and time organisation. Meanwhile the teachers' union has been increasing concern that the move towards school autonomy is, in fact, detrimental at a time of financial constraints and restrictions.

The 14th School Organisation Act defines the general conditions in the law and thus has initiated a new philosophy of school development: Possibilities for school autonomy have been created for all schools. The extent and the form of the autonomy have been differentiated according to the various types of schools. The regulation of details that would normally have resulted in nationwide school conformity was waived; this has paved the way for more school-based management and, as a consequence, evaluation becomes more important and more or less a duty for everybody within the framework of school autonomy. The first research results about the consequences of this development have shown that schools react differently and can be classified among the following three types (cf. Krainz-Dürr et al., 1997):

- **Type A schools**: These are innovative schools throughout the country which had not waited for the authorities to ask them to develop their own school plans. Years ago they had started analysing their status quo and have continued re-organizing and evaluating school life ever since. The new laws concerning school autonomy might have made things easier for these schools, but they are yet ahead of national developments.

- **Type B schools**: A great number of schools and a majority of teachers have not yet been convinced of the necessity of self-evaluation; for them evaluation is still restricted to giving students marks on tests and being assessed by their superiors. These schools cannot easily accept the idea that it is necessary to regularly evaluate their schools or that parents and students can have a say in matters of school quality.

- **Type C schools**: After hesitating for quite a long period an increasing number of schools have started to make use of their newly achieved educational freedom searching for solutions to their particular problems at the local level: New subject areas are being introduced, informal networks are being created and a sense of change is noticeable. They have started to look also into their weak areas rather than presenting the 'sunny side of their facades'.

Making use of this growing awareness, the ministry of education has made the evaluation of quality a major political theme, which should lead to:

- a step-by-step approach towards mandatory self-evaluation at the school level within the next few years;
Self-Evaluation as Endeavour to Understand Quality

• a new structure concerning the system of school inspection (cf. Schratz, 1998);
• measures in the areas of crisis intervention and system-wide evaluation.

Several measures have been introduced to start implementing the political framework by utilising various experiments and experiences in relevant areas so as to work towards a country-wide implementation of the overall policy regarding quality development within the Austrian school system. Moreover, those activities should also contribute towards building up a necessary support system for the implementation process. For example, in the area of self-evaluation a quality network has been developed which monitors different approaches (e.g. TQM [Total Quality Management], EFQM [European Foundation for Quality Management] etc.) in the area of occupationally oriented schools, within a regional school network the FQS (Formatives Qualitäts-evaluations-System) has been tried out (cf. Strittmatter, 1999), as well as other initiatives have sprung up on the regional, national and international level. When the European Union introduced the pilot project “Evaluating quality in school education” during the school year 1997/1998 (cf. MacBeath et al., 2000), the ministry used the project’s impetus as a springboard for establishing a nation-wide self-evaluation programme called Q. I. S. (Qualität in Schulen [Quality in Schools]). It comprises a support system consisting of back-up materials, school networks, regional workshops and expert advice helping schools to help themselves in school-based evaluation. (The latest developments can be found on its website under www.qis.at.)

SELF-EVALUATION WITHIN THE EUROPEAN EXPERIENCE

The pilot project was the first large scale European effort in the school sector in the field of quality evaluation. Its objectives were:

• to raise awareness about the need to evaluate secondary education in Europe;
• to enhance existing national procedures;
• to give a European dimension to quality evaluation;
• to support the exchange of information and experiences.

On the the local level it should help schools in introducing self-evaluation as a major quality development philosophy including all stakeholders and in implementing measures of school development as a result of the findings. Five Austrian schools participated in this European project to test new evaluation
methods of and develop strategies for the evaluation of quality in school education.

Discussions on what quality of school education actually means has become the starting point for all schools. Although many research results from school effectiveness studies have pointed towards success indicators of what could be regarded as "good" schools, long lists of indicators would have estranged the participants in the project to find out themselves about what quality actually means for them and, with regard to the overall system, for schooling in Austria at large. To do so, the schools used a tool developed by the steering group of the project (cf. MacBeath, Meuret & Schratz, 1997) called the Self-Evaluation Profile (SEP), which covers the following areas (see Fig. 1).

The schools participating in the project were presented with the SEP (see annex) and – taking the twelve areas into consideration – were set the task of compiling a picture of the quality of their schools from various perspectives, and of ascertaining to what degree the opinions of the individual ‘actors’ were compatible with each other. Using these results as a basis, the schools then decided upon the aspects they wished to regard more closely – and thus to submit to a more detailed evaluation. Each school was to concentrate on five areas in order to enable a wide spectrum of characteristics of quality to be recorded, on the one hand, and not to get caught up with long-winded enquiries, on the other hand. The schools participating in the project were given a free hand as to the choice of a suitable method of evaluation. However, they were expected to apply acceptable methods and underlying principles of evaluation.

Fig. 1. Areas for Self-Evaluation.
in their school-based evaluation efforts. They were also expected to participate in an exchange of ideas and information with the other schools taking part in the project.

This exchange of ideas took place at various levels:

- **The individual school:** A project group was formed at each school. It was responsible for the preparation, execution and evaluation of the measures to be taken in the pilot project. Desirably, students and parents were to be included as members of this group, if possible. In addition, each school chose a 'critical friend', who would give an outside, un-biased opinion when necessary (cf. MacBeath, 1998). The critical friends were chosen by the schools, themselves. They were (among other things) representatives of different professional backgrounds (e.g. further education, the sciences and the business world).

- **The national level:** A National Committee was formed in each country participating in the project. It was responsible not only for the co-ordination among the individual schools, but it also secured a link between the European Commission and the political educational development of the country involved. Exchange meetings among the five schools in Austria were designed not only for exchanging ideas and experiences, during the two-day meetings, but also for developing further concepts of school-based evaluation.

- **The transnational level:** Individual schools made contact with schools in other countries of the European Union which were participating in the project. They exchanged information and experiences and became familiar with specific aspects of the project. This contact took place not only directly between the schools but one also communicated via computer e-mail and held video conferences.

  Apart from this, there was also a transnational exchange of schools of various countries. For example, in February 1998, members of project schools from Germany, Liechtenstein, Luxembourg and Austria met in a teachers' in-service training institution in Germany for a workshop, in the course of which experiences were discussed and schools were visited.

- **The European level:** In November 1997, the schools participating in the project of the European Union met for the first time in Luxembourg. The way was paved for further exchange of experiences. A Newsletter was compiled as a written communication and an internet conference was also set up among the schools. Furthermore, a home-page (http://europa.eu.int/en/comm/dg22/poledu/indb-en.html) offered comprehensive information about the status of activities in the current project, and links provided information about the individual schools and their various activities. In the autumn of
1998, during Austria's presidency of the European Commission, a final meeting was organized in Vienna.

**SELF-EVALUATION AT WORK: SCHOOL-BASED EXPERIENCES**

Rather than describing from a distant research perspective how the participating schools took the road towards self-evaluation, we asked the schools to give an insight into how they experienced key aspects during the self-evaluation work. During mutual school visits the school teams, who also observed classroom teaching and interviewed some of the actors involved, helped to validate the data. Thus the following accounts reflect multiple evaluation perspectives from the schools involved in the project.

*Case 1*

At the grammar school BRG 19 in Vienna several working groups were concerned with the personal and social development of students, educational difficulties, time consumed in the learning processes, the satisfaction of teachers and students with their working conditions and the quality of co-operation between parental home and school. The suitability of the self-evaluation profile (see annex) as an initial impulse in striving for a mutual goal, can be substantiated by the following account.

7th November, 1997 – The arts and crafts room of the above-named grammar school: A confused din of voices; expectant faces; teachers, students, mothers and fathers zealously discussing the purpose of the meeting. The key-word is 'evaluation'. A so-called 'self-evaluation profile' of the school is to be compiled. There is a good deal of interest, especially amongst the teaching staff. One becomes inquisitive and attentive as the head of the school begins to explain the purpose of the meeting and the procedure to be followed. Coloured slips of paper are distributed and six groups are formed according to the colours. Each group consists of two students, two parents and two teachers. The moderators – two students, two parents' representatives and two teachers – get to work with their groups.

For the first time, the members of the groups are presented with the self-assessment profile with its twelve criteria of quality. Certain definitions are explained to the groups, but then each individual member must analyse the profile in his/her own way; an evaluation of the current position and an
evaluation of the development of the school in each area over recent years must be made. One can sense the tense atmosphere — such as exists during examinations — one racks one’s brains, one considers one’s answers carefully — and, naturally, there is no cheating! Everyone makes a great effort — not only the students!

Of course, this is only the beginning. After this individual phase, things really get moving: a consensus is sought within the individual groups. The ice is soon broken and the moderators are kept busy. The fact that four of them are students or parents makes the work even more stimulating. It is certainly not an every-day constellation, but it functions — they do their jobs well and are accepted by the other participants. It is certainly no easy matter, if one considers that occasionally a teacher must be reminded by a student to keep to the set time-limit when expressing his/her ideas . . . ! It soon appears obvious that a group consensus can easily be achieved in several areas; a few points call for more extensive discussions and, in a few cases, one can only agree to disagree and accept only two evaluations as the opinions are far too diverging. A certain lack of clarity appears even in very fundamental questions ("What’s meant by ‘personal and social development’?") and the four-stage evaluation scale, which allows no room for retreat to a neutral position, ensures that the discussion remains lively.

Much later, one moves on to the third stage of the project. It all sounds so simple, but actually the participants must gather their last resources of energy to compile the final agreement of consensus of the entire group. No easy task if one considers that each of the (voluntary) speakers of each group must represent the ‘consolidated’ opinion of his or her group. After lengthy, tedious, animated discussions, the participants finally leave the school building. It is late evening, but they sense a feeling of satisfaction. They have worked out the foundations for further participatory action concerning evaluation in school education. What is more, one has taken time to talk to each other, to discuss mutual problems in a relaxed atmosphere and in a candid manner. Partners in school affairs listened to each other, gave each other the impression of playing an important role in striving for a common goal. Perhaps a few of them returned home with the feeling of having seriously touched upon certain topics for the very first time.

Reactions on the ‘day after’ were positive. Parents and students apparently felt that they had been accepted at the meeting as active partners in school affairs. Since then, the development has become extremely self-potent; new points of view have arisen and further aspects have been conceived. A critical examination of quality and its development is well on the way to becoming a fixed component of our school culture!
Case 2

The secondary modern school in Gries at the Brenner Pass decided to use a standardised questionnaire which had been developed to evaluate the school climate in Austrian schools. One group investigated the delicate question as to whether, and to what degree, the present school organisation promotes the personal and social development of students. A further point of interest was the proficiency of the students in mathematics in comparison to other Austrian schools throughout the country. The school, as a whole, showed great self-confidence in expressing its willingness to be involved in a comparative evaluation of the school climate.

Of course, not everyone would agree to lay himself or herself open to such an examination – after all, the results might prove unpleasant! However, personally, we did not regard the official survey as a threat. Our pre-requisites and feed-back were good – we could expect a confirmation of our strength rather than a revelation of our weaknesses. Our school is attended by students of all the elementary schools of the district. On an average, there are 21 students in a class. We have only eight classes, so that everything is clearly arranged. The structure of the students' family backgrounds is, for the most part, conducive to education. The selection of rooms and the equipment of the school can meet the most pretentious claims. The time-table of lessons is set to co-ordinate precisely with the bus time-table – a very significant mark of quality in this region. Our students have only 30 compulsory lessons a week and ample optional subjects from which to choose. We offer team-teaching; group-teaching in small groups; social studies in the first form; regular meetings of form representatives; a comparison of gradings within the school and an internal agreement to allot marks only from 1–3 in subjects such as arts and crafts. We are also traditionally renowned for respecting the needs of the less-gifted students – so we had nothing to fear. We did not expect 'bad marks'!

Nevertheless, we regarded the official survey and its evaluation with expectation and great interest. Firstly, naturally, we wished to have our presumptions confirmed and to be measured on an Austrian scale. Secondly, we wished to experience, personally, how such professional evaluations are carried out. Naturally, we were also curious about the results of the differences within the individual classes and in which areas we would prevail (or otherwise!).

After our critical friend had explained the purpose of evaluation to the members of staff in a very competent and trustworthy manner, and after everyone could be assured that the survey was not designed to aim at anyone...
in particular and that only a small ‘evaluation group’ would be involved in extra tasks, the attitude of the teachers towards the survey was favourable or, at the worst, neutral.

The critical friend, herself, who is a university researcher, was solely responsible for carrying out the survey. Retrospectively, we realise how important that was, as she had ample experience in dealing with such surveys and, moreover, as an ‘outsider’ could guarantee a candid, unbiased feed-back.

In the first forms, the survey lasted approx. 60 minutes; in the fourth forms, 40 minutes. The questionnaires, which were filled in anonymously, were sealed and posted to the research centre to be analysed.

In February, the results arrived: 54 pages long – one school result, eight form results, each in 14 different areas and a precise presentation of the 65 questions – complete with comments. We realized immediately that the students and parents should be presented with the overall results. At the same time, it became clear, however, that only the teachers, themselves, should be allowed to ponder over the results of the individual classes. We did not wish to provoke unfair, contra-productive comparisons among the classes by presenting these results to the students and parents.

On casting a first glance at the graphical presentation, we were immediately relieved: we were above the Austrian average in 13 out of 14 areas. Regarding ‘proficiency drive’, we were given a mark of 91, which is 9% below the Austrian average, and for ‘fairness’, we were allotted 107 points – 7% above average.

Then followed the exciting comparison of the classes: We were aware of the fact that they differed. But to what extent, and in which particular fields, would this become apparent in the evaluation? First of all, the head of school invited each individual form teacher to a discussion about ‘their’ forms’ results. These discussions usually followed the same pattern: one commenced with an estimated comparison of form/school results; continued with the enlightening experience of one’s mistakes suddenly becoming evident; one then attempted to justify one’s actions and quite often finished up in a state of pensive meditation. It must be noted, that in certain areas, there were discrepancies of up to 16% points among individual classes! Hasty interpretations showed that boys, particularly, felt at a disadvantage when dealing with male teachers and that a form with certain ‘trouble-makers’ regarded itself as being put under pressure. Self-critical considerations also came to the fore – could a decrease in proficiency drive also hint at a decrease in proficiency? Perhaps our concern in this respect is the reason for our planning to make efforts to obtain objective feed-back concerning school proficiency as soon as possible? This way of thinking has just taken root and the results are just as incalculable as our reactions towards other results of the survey.
What has the survey proved so far? The general opinion is that such surveys are not empty fallacies; that 'gentle' evaluation is no problem; that they arouse awareness and bring certain facts to light; they stimulate reflection, urge one to put things into perspective, influence the sub-conscious mind and might even lead to a few new resolutions, corrections of behaviour and simple improvements in quality being made. Perhaps such surveys also stir curiosity and awake an urge to investigate further – thus preparing the way for a culture of evaluation to be secured in the future.

Case 3

A secondary school in Upper Austria, Hauptschule Pabneukirchen, chose a similar context of questions, but a completely different methodical approach. This school dealt with the students' proficiency in mathematics according to chosen items taken from TIMSS (Third International Mathematics and Science Studies). They also held teachers conferences on the topic of school development, carried out surveys on the burden of homework, applied focussed analyses (e.g. on the use of atlases in geography), but in order to find out more about the atmosphere of the school, they chose photo evaluation as a method (cf. Schratz/Steiner-Löffler, 1998).

In this method of evaluation, the students, themselves, estimate the school as a place of learning and communal living. The students take photos of places at school where they feel happy and at ease and places where they feel decidedly uncomfortable. The photos are presented on posters with short descriptive texts. The students state why they feel particularly at ease in certain places and what could be improved in the 'unpleasant' places at school.

When planning the project, great care was taken to ensure that at least one group of each school-year was included in the activities, as it is more than likely that students of different ages hold different opinions regarding the topic 'feeling happy'. Each group, which consisted of approximately five students, chose four locations where they felt comfortable and a further four locations where they felt decidedly uncomfortable. Then the students considered the point they wished to stress about each object and how this could best be represented by a photo (e.g. by sitting very closely together in a 'cold' room). Finally, the respective objects were photographed with an instant-photo camera, the photos were mounted on a poster and a short text was written and graphically supported (e.g. a 'happy' place - sunshine; an 'unpleasant' place - rain clouds). When the posters had been completed, they were presented – with a short
Self-Evaluation as Endeavour to Understand Quality

explanation - to the other students who were participating in the project and finally discussed.

During the course of the project, 12 posters were produced. They were hung in the main hall of the school for eight weeks. In this way, the other students, who had not been able to participate in the project, were given the opportunity of being confronted with the ideas and purpose of the photo evaluation. The students were extremely industrious. For example, they worked through break without the least complaint! They enjoyed being allowed to express their feelings through photos. It was evident that the less-fluent students were particularly active! Using this method of evaluation, they were able to express their opinions in a visual manner (photography). Compared to the usual methods of self-evaluation in the written form, in which students who have a fluent command and cognitive comprehension of the language can express themselves more easily, photo evaluation enables a more comprehensive observation of the school!

The discussion process at student level is being summed up at the present time. The head girl of the school invited the form representatives of each form to name the three most important wishes concerning improvements of the school building. In this way, a students' 'Catalogue of Requests' was compiled – 24 requests, which the head girl then handed over to the school management. The students were also allowed to present their requests to a teachers' conference. A joint discussion of students and teachers representatives then decided which requests could be met within a given time. Did requests involve building measures, the school authorities, naturally, had to be consulted.

The school management and the teachers involved were naturally fully aware of the fact that, by encouraging this photo-evaluation project, they had indicated their willingness to carry out alterations in the school building. Expectation stirred among the students that the teachers would also participate in making the school more attractive. A few of the requests which were expressed in the photo-evaluation could be met with little cost and effort and have already been completed in a joint effort of teachers and students (e.g. arranging and decorating the classrooms). It is of particular importance – if some request or other is rejected – to explain to the students the justification of the decision in each particular case. Reactions and discussions among the staff towards certain decisions which had to be made, could prove to be a further positive step towards the development process of our school, if managed wisely.

Case 4

"Actually, we are (not) very well-known!" is the title of a report of a survey undertaken by a secondary school in Graz, Neue Mittelschule Puntigam.
Undertaking this survey, the school attempted to gain a picture of just how well-known it actually appeared in its immediate surroundings and to what degree it was accepted. Various studies ran collaterally or in collaboration: 'Disturbances during lessons' was the central issue of a topic tackled by guest students and teachers; the educational conditions and the degree to which the special needs of particularly talented and less gifted students are met, were topics of other areas. Public opinions, interviews and photo-evaluation (see above) were also applied in this project.

In the second form of this secondary school in Puntigam, we decided to encourage the students to carry out a project during the German lessons on the topic 'School and Community'. For some time now – and especially considering the fact that in the present academic year we only had sufficient students for two first forms, instead of the three which had been planned – we tended to believe that although there were noteworthy activities within the school itself, there existed a certain deficit as far as publicity and public relations were concerned. We therefore decided to make a clean start and check-up on our popularity. How well-known are we? How do other people regard us? What have they heard about us? With our contribution, we hope not only to serve public relations, but also to support the further search of identity and the self-assurance of the school.

The Phase of Preparation During Lessons
The introduction of the project is very promising. The students need little animation to accept our plans – they participate with glowing ardour! Our first task is to list the various institutions, public facilities, establishments etc. (infra-structure) of the region Puntigam and possible interview partners. The lists are noted on the blackboard. Seven groups become apparent out of the thematic summary and structure: public facilities (nursery schools, day nurseries, churches, post-office); medical facilities (doctors, chemist's); public services (cleaner's, travel agents, solarium); guest houses, shops, political representatives, pedestrians. The students form groups of three or four and consider suitable questions to be posed in interviews in their particular area. These questions are then compared, co-ordinated and compiled into a list of 12 suitable questions to be used as a guide by all the groups. In the final stage of this phase, the students take on 'roles' – four 'roles' per group – and we act out the interviews. The roles are: 'the person who makes the first contact with the person to be interviewed'; 'the person who explains the
Self-Evaluation as Endeavour to Understand Quality

project to the person to be interviewed'; 'reporter' and 'technician'. Need we mention, at this point, that this method of acting out a situation, not only serves the purpose of relating to reality, but is also great fun!

Interviews and Evaluation
The 22 students get to work, armed with dictaphones, note-pads and the headmaster's official written approval. They carry out 25 interviews in all corners of the surrounding district of the school. Next day, at school, there is a general discussion in the class of the experiences gained. The teachers are particularly keen to hear what the children have to say. Each group has something of interest – or an amusing episode – or something amazing – to report. The most difficult task appears to have been set for the group of students designed to interview pedestrians. The girls were not to be put off, however, by the constant rejection shown to them. They did not despair and completed the task they had been set. Putting the interviews to paper proves to be a very tedious task, although we do not make much of a written transcript and encourage the students to concentrate on the essential points of the answers. The students, themselves, evaluate the results of the interviews. They present their findings in a comprehensive chart. The teachers work out the percentages, as we regard this task to be too exacting for the students!

Results
We are, naturally, fully aware of the fact that we cannot speak of representative results, since we only questioned 25 persons. Nevertheless, however, presumptions can be drawn from the data we obtained. We are also of the opinion that it would be profitable for the school to follow up a few of these facts. Some examples:

• More than 60% of the people interviewed knew that we practise team-teaching; that we have three choices of school-leaving certificates and that one has the opportunity of learning a second foreign language from the third form onwards;
• 56% were aware of compulsory project teaching in the school but only 40% had heard of 'social studies'. The various extra-curricular, out-of-school activities were approved of and supported by all the people interviewed;
• Approx. 50% of the people interviewed could name the headmaster or at least one teacher;
• Our type of school was generally accepted (76%), only one person rejected it completely. 20% of the answers had to be set aside, as they were irrelevant to the questions;
• 16% of the people questioned in the district knew nothing of our school;
• Only 44% had heard of the term 'Schulverbund West'. (This school is one of five schools in the western region of Graz which have united to an interlocked system).

What have we Learned from the Project?
As far as lessons were concerned, the students were particularly motivated by the unconventional approach. The phase of preparation was, therefore, delightful and the children showed great commitment. The students' understanding for the necessity of practising the 'roles' and of compiling a 'professional' interview guide (PC) proved to be a pleasant change in everyday lessons for all concerned. We – the teachers – realised that our students – for the main part – were capable of and willing to make appointments, to organise the necessary equipment and to carry out interviews independently. Initially, we under-estimated the twelve-year-olds in their proficiency in the preparation phase – we never once doubted, however, that they had the courage and spirit of adventure to tackle the interviews! Our strong support was called for, however, in compiling the written results to ensure a minimal degree of comparison. We found it remarkable that not even the children, themselves, were interested in the overall results – only very few showed any interest at all! ('Let bygones be bygones')!

How relevant was the study and the results for our school in general? For us, the most striking answer to the question, “Do you know anything about our school?”, which was posed by the students in a bank located in close proximity to the school, was a simple, straightforward – “No!” This curt answer, although it cannot be regarded as representative, nevertheless, led us to believe that, in future, we must concentrate more on public relations and publicity; that we must project information about our school and its aims to the community outside school and offer more transparency for the public into our internal affairs. Considering the results of the study, it would appear extremely meaningful and necessary to carry out a precise analysis of the strategies of passing on information of our school. The first step in this direction has already been taken by the interviews we have made. We were also amazed by the fact that, in spite of several friendly chats with colleagues around the staff-room coffee table, only a few were really aware of our undertaking. The two of us actually had the feeling of acting in empty space – even if – ironically enough – the topic at hand was, ‘School and its Surroundings’! Were we offering too little transparency? Was there a lack of interest in the topic? Or were several colleagues simply shunning extra work? We do not intend to disregard these
unanswered questions. There would be no point in doing so – neither for ourselves nor for the school.

Case 5

The fifth Austrian school to participate in the project, the Gymnasium in Stainach, had already carried out a very comprehensive, expensive analysis of its current situation using the model TQMS, and had thoroughly analysed the results. At that time, they have arrived at the stage of compiling a guide and of putting the knowledge they had gained into practice. However, this did not hinder the participants in any way from continuing to strive for further improvements – for better ways of presenting lessons; for new conditions in the classroom; or for possibilities of improving the atmosphere at school even further. The analysis made it possible to identify the need for necessary improvements and to draw a self-evaluation profile of the school. As a consequence, one became occupied not only with advisory measures in the fields of contact between elementary schools and secondary schools, but also with an ‘early warning system’.

A student asks a teacher, “Excuse me, Miss. Do you sell the cheese sandwiches?”

That is no fairy-tale! Nor is the following observation: the student who courageously drags complete works of every possible text book available and piles of exercise books to his new secondary school every morning... Such observations caused us to take the initiative whose aim it is to simplify the transference of students from elementary schools to secondary school. The overall programme consisted of three measures:

Open Day:
Approximately three weeks before the final date for application for registration, children who have completed elementary school are invited to view the secondary school with their parents and to take part in lessons of the first two forms.

Bridging the gap between elementary school and secondary school:
Every year, towards the end of March, the grammar school invites all the teachers of the final year of the elementary schools in the district to discussions in small groups with the teachers of the grammar school who will be in charge of the children in their first year there. The results of these discussions are presented to the whole group, in plenum. On the one hand, they offer valuable information
about the level of proficiency of the children leaving elementary school and, on the other hand, the expected requirements of the young grammar school children are clearly stated. Occasionally, the teachers even attend each other’s lessons. Institutionalisation of such peer observations would be desirable in this respect.

Project: A Tiger’s Time-Table:
Students of the 11th grade who have chosen ‘Psychology, Philosophy and Education’ as an optional subject have been intensively occupied with educational psychology in theory and practice for approx. six weeks. Together, they have compiled a script (which features a very strong tiger on the cover page) for their young school comrades. On two afternoons, they offer to help the young students to organise their time-tables – especially as far as homework is concerned. These students, who act in an advisory capacity, work in pairs with 5–7 ‘clients’. Participation is free and voluntary. On an average, approx. 75% of the first-form students take advantage of this offer. We are considering extending the project towards a regular service during the entire first term.

“Help! My child is in trouble!”
§19. Par. 4 of the School Education Act states as follows: “If a student’s proficiency furnishes sufficient proof to worthy its being graded as ‘unsatisfactory’ in a compulsory subject in the second term, the parents or other authorised persons should be informed immediately and, together with the students, be given an opportunity to discuss the matter with the form teacher or other teachers of special subjects. At the same time, measures should be taken... to promote proficiency and thus to prevent a negative grading.”

This ‘Early Warning System’ is designed to signal a new culture of passing on information in schools. Its characteristic is that parents, or other authorised persons, are informed on time and useful, necessary information is passed on. I would like to give a short description of how this new system – ‘Advisory Counselling instead of Negative Gradings’ – actually works at our school.

Organisation and Timing:
Although a specific time limit is no longer set for determining which children are in danger of being negatively graded, our school regards the end of April as a suitable date. The teachers name the students in a list of the appropriate form and a class conference is held in which the form teacher and the other teachers of the affected students discuss the situation – keeping the guide for advisory counselling in mind. If a student is affected in more than one subject, then it appears to us to be of particular significance, that the individual teachers
agree upon a pre-determined procedure. The second parent-teacher meeting takes place at the beginning of May. The invitation which is sent out indicates that the main topic to be discussed is the 'early-warning system'. To our mind, this consultation should be attended particularly by parents of students who are in danger of being negatively graded in one or more subjects. Such advisory counselling (parents-teachers-students) is designed, moreover, to prevent information from becoming mislaid and other such misunderstandings.

Course of Procedure:
Both the class conference and the advisory counselling could run according to the following pattern:

- **Current situation – description of problem and analysis of problem:** How and why have deficiencies in learning and knowledge arisen?
- **Desired situation – listing of aims and possible forms of support:** What can be done and how?
- **Declaration of ultimate aims – measures and concrete steps.** Who will do what and when?

Evaluation:
After a certain length of time has passed (when the grade of proficiency has been determined) or at the end of term (when a positive or negative grading has been decided upon), it would be of value to check on whether the agreement of measures to be taken has been kept and on the effectiveness of such measures. This discussion can also take place in the absence of the parents, but is, in any case, voluntary.

Centralised, transparent methods within the school itself, close co-operation of all the teachers of a class among themselves and with the form teachers, as well as suitably structured advisory counselling among the teachers, parents and students – all these things contribute towards a new school culture which is focussed on the success and well-being of each individual student.

**CONCLUSIONS FROM SCHOOLS’ EXPERIENCE IN SCHOOL-BASED EVALUATION**

School development – of which quality evaluation is a component – takes time. Anyone who knows the rhythm of such processes is well aware that it is much too early to draw extensive conclusions from such experiences. It would appear
more appropriate, in this connection, to speak of initial contours and assumptions – which are, indeed, most promising.

Feed-back from the schools, and observation of occurrences there, hint at one particular fact: There is, quite obviously, an increasing need for quality evaluation in the school system – and schools are becoming increasingly conscious of this need (cf. type C schools above). Considering the basic respect one holds for the accountability of schools and the lack of experience at all levels of the school system as far as safe-guarding and developing quality are concerned, the significance of approaching this complex task cautiously and of suitably preparing the necessary framework of regulations, becomes apparent.

The 'protagonists' among the teachers, parents and students, although armed with prudence and enthusiasm, are, however, confronted with a framework of conditions which threatens to influence not only the onset of such processes but, particularly, their lastingness. The Austrian school system lacks not only a culture of systematic evaluation and the necessary training for student teachers, but the schools, themselves, also lack the all-round internal structures which are necessary to support such processes in the long run. Once again, it becomes clear just how imperative it is to discuss and reform the framework of conditions for school development (the task of teaching, working hours, resources, managerial structures, new models of in-service training etc.) as widely as possible including as many 'actors' as possible.

As far as the initial onset of the quality-evaluation process is concerned and how it is put into practice, it seems to have proved useful, at the start – when summing up the situation – to cast a very wide-meshed net rather than to become involved with intricate details. Thus, the self-evaluation profile, which was specially developed for this project, appears to be a very suitable instrument at the initial stage. Apparently, precision is not particularly decisive during the first stage, but rather the motivation and potency which develop out of the communication among the participants – something which is often experienced as a completely new quality. This is particularly true of schools which have 'dared' to seriously incorporate students and parents into the efforts towards evaluation from the very start. Incorporating an outside opinion in the form of a 'critical friend', or encouraging a mutual exchange of school visits, evidently induce not only new perspectives to be opened up, but also increase the motivation of the participants. During the course of the evaluation, one might be concerned with seeking a suitable balance between the seriousness and profundness of the study, on the one hand, and its lastingness and feasibility, on the other. Original methods and instruments of evaluation – for example, photo-evaluation can prove useful in this context. Naturally, not even such methods
can assure the success of the initiative if the purpose of the activities is not clearly stated from the very start.

On a more general level the experiences in school-based evaluation have thrown new light on the steering forces in educational management (cf. Schratz, 1999a). Every endeavor in evaluation and quality development is grounded in an intricate field of tension characterised by the dynamic interplay of (at least) three dimensions, which can be depicted in the shape of a cube (see Fig. 2 from Schratz, 1999b; see also MacBeath, 1999, pp. 2–3).

The first dimension arises from the relationship between bottom up (†) and top down (‖), that is between self-evaluation and evaluation (intended) by others. This is a matter of ownership or, in other words, who is in control. Does the school want to find out about certain issues or is it somebody else’s idea (e.g. superiors, school board etc.). Accordingly, a bottom-up movement is when the school in Pabneukirchen wants to find out about the students math proficiency (see example above), an international study (like TIMSS) using a standardised test instrument to find out and compare students’ achievement is a top-down movement. The difference can be shown in Fig. 3.

In the bottom-up approach the evaluators want to find out about the effectiveness of certain programs or actions (cause and effect). Here evaluation is the endeavour to understand the quality of processes and products. In the top-down approach the evaluators want to get a clear result according to given standards or norms so as to compare and contrast findings from different samples. Here evaluation is the assessment of how certain norms are fulfilled.

The second dimension arises from the relationship between internal and external evaluation (←→). They refer to whether the school’s own means (including human resources) are sufficient for the evaluation or if somebody from outside is contracted such as a ‘critical friend’ in the project described above. This principle is important in peer-visits in order to bring in the outside view and thus to throw light on the ‘blind spots’ of the insidedness with a view to enhancing quality discussions.

The third dimension refers to the pressure-support axis between the push (∫) and pull (∫) forces in a development process. At one extreme there is full support by the system which allows schools to freely develop into the direction of their desire, at the other end there is extreme pressure by rigid accountability. A balance between those extremities seems to make best use of intrinsic and extrinsic motivation for schools to exceed expectations.

The findings of school-based evaluation approaches in Austria have so far shown that the ‘right’ balance between the extremities in the ‘Evaluation Cube’ (Fig. 2) is critical to the success of quality development. If there is a one-sided government initiative implemented only from ‘above’ it will fail a
Fig. 2. Dimensions of School Evaluation and Development.

Fig. 3. Evaluation Between Activation and Comparability.
large-scale identification by the key players in the field. Internal evaluation has to be met by some external body so as to be linked up with the larger system. If there is too much weight on the external evaluation (e.g. a rigid inspection system) schools will not feel responsible for quality themselves. If the state simply puts pressure on schools to evaluate themselves but does not offer a support system enabling them to do so, the results will suffer again.

"Achieving the [optimal] blend is the key factor in determining whether schools will grow and flourish or stagnate and decline." (MacBeath, 1999, p. 3) Quite agreed! The challenge, however, is to satisfy the different needs and expectations from above and below, internally and externally in a situation where society seems to widen more and more and school systems have to scope with a plurality rarely experienced previously.

**NOTES**

1. Edwin Radnitzky from the Austrian Ministry of Education, national coordinator for the Austrian schools, and Michael Schratz, scientific advisor for the Austrian schools and member of the steering group of the European Commission.

2. I thank Claudia Kreulitsch, Monika Pantlitschko and Susanne Wiblinger (BG/BRG 19, Vienna), Hubert Jungwirth, Brigitte Renzler and Helga Schwarz (HS Gries am Brenner), Thomas Zwicker (HS Pabneukirchen), Gertraud Grimm and Ingrid Ninaus (Neue Mittelschule Puntigam) and Maria Haindl (BG/BRG Stainach) for the accounts they have contributed. They originally appeared in *Erziehung und Unterricht* (1998) 148(7/8).

**REFERENCES**


**SELF-EVALUATION PROFILE**

Please evaluate the school’s standing in relation to each area on the following scale:

<table>
<thead>
<tr>
<th>Area</th>
<th>Rating</th>
<th>Recent Evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Academic achievement</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>2 Personal and social development</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>3 Pupils destinations</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td><strong>Processes at classroom level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Time as a resource for learning</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>5 Quality of learning and teaching</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>6 Support for learning difficulties</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td><strong>Processes at school level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 School as a learning place</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>8 School as a social place</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>9 School as a professional place</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td><strong>School and environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 School and home</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>11 School and community</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>12 School and work</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td>++</td>
<td>+</td>
</tr>
</tbody>
</table>

1 Please refer to the description of the areas in the guidelines (MacBeath et al., 2000, pp. 102-114).
INTERNAL AND EXTERNAL SCHOOL EVALUATION IN THE SPANISH EDUCATIONAL REFORM

Alejandro Tiana

INTRODUCTION

In 1990, a new education act (known as the LOGSE) introduced a series of far-reaching changes into Spanish education. These changes included the structure of the education system, the process of curriculum design and development and the degree of autonomy granted to schools in organising their activities. This new law marked the culmination of a reform process which had been underway since the nineteen eighties. The changes introduced entailed a relatively in-depth reorganisation of both individual schools and the education system as a whole.

THE RATIONALE FOR SCHOOL-BASED EVALUATION IN THE SPANISH EDUCATIONAL REFORM

One of the features introduced by the 1990 law was the creation of a new body whose task was the overall evaluation of the education system. In order to fulfil its remit, the INCE (Spanish National Institute for Quality and Evaluation), as it is known, needed the cooperation of the regions (or "Autonomous Communities") into which Spain is divided and which have a wide range of powers on education and training. The new Institute quickly began its work.
designing and implementing nationwide studies which focused on the reform process underway.

In Spring 1996, the INCE published its first report assessing the implementation process and outcomes of the new Primary Education. Covering a wide spectrum of different variables, this ambitious study was conducted at the time when this new educational level was being introduced. One of the main but not sole aims of the study was to assess pupils' outcomes with a view to drawing up a base line from which progress could be assessed over time.

A large number of data were published by the Institute itself in two reports of an essentially descriptive nature (INCE, 1996, 1997). However, other data of a more analytical and inferential kind, attempting to explain some of the findings were not published, despite having been circulated by the press, being included in an unpublished version and having been available for a while on the Internet. Some of these were of considerable interest in providing a clearer picture of the Spanish education system and its operation.

One of the issues that most interested the authors of the study was a comparison of the results obtained by pupils in public schools with those from the private sector. There is a widespread belief in Spanish society that private education achieves better results than the public sector, a belief which warranted a closer look in the light of objective data. The overall findings appeared, in principle, to justify this opinion, since in the different school years and subjects evaluated, private school pupils indeed obtained better scores than their counterparts in the public sector, the gap of about ten percentage points between the two being a relatively stable one. However, on closer analysis and bearing in mind the socio-cultural origin of the respective pupils, it soon became clear that this difference was due not so much to the type of school as to the students' social background. When the scores were weighted to counteract the effect of this socio-cultural origin, the results were actually found to be very similar and the differences between them proved to be of little significance.

Of course, there is nothing new or original about this observation. In the nineteen seventies, the famous Coleman report fuelled a major debate in the United States and other countries on the influence of pupils' social background on their academic performance (Coleman et al., 1966). This type of study was to provide the basis for the lines along which research is still conducted today and which tends to produce broadly similar results. Today, educational sociology therefore gives general credence to the significant influence of social background on pupils’ results at the primary level, although there is more disagreement about how far this influence applies to higher educational levels.

The results of the INCE study completely coincide with international research findings, which is no particular surprise. However, as was also true for the
Coleman report in its day, these data could appear to back the idea that school is no more than a mere reflection of social reality and that pupils' results are determined exclusively by their socio-cultural origin. In fact, most current analysis goes some way towards relativising this markedly deterministic theory. Whilst the reproductive role played by schools in society today cannot be denied, nor should its differential effect – both in social and individual terms – be underrated.

The existence of such a differential effect is made patent in a range of studies, including the INCE research analysed here. Although countering the effect produced by social-cultural background (by the statistical calculation of a specific index) caused an obvious levelling in the results when comparing public schools against private, the same did not occur when an identical operation was conducted with the results obtained by pupils in different regions or when comparing similar schools. In the first case, the differences between the regions were sometimes lessened, but at other times they proved even starker than before. In the second case, on the other hand, schools with a very similar kind of pupil nevertheless obtained results which diverged considerably from one another.

The first phenomenon can be accounted for by the influence exerted by factors such as diversity in education policy, social expectations and attitudes towards education or the organisation and operation of the education system in a particular region. It is worth pointing out that many of the differences noted between countries in international evaluation studies, such as those undertaken by the International Association for the Evaluation of Educational Achievement (IEA), can be explained by this type of factor. This is an area where major differences arise that are related to education policy and administration.

The second phenomenon involves the operating style of schools and the way in which they organise their educational work. Indeed, schools exert an educational influence on their pupils which goes beyond what might strictly be expected on the basis of their sociological characteristics. If pupils' social background were the only variable determining performance, we would not expect to see major differences between schools serving the same population sector. And yet the data confirm that there are indeed marked differences in performance from one school to another.

It is precisely this last observation which has served as grist for the mill in the different schools of thought emphasising what is known as the school effect, a phenomenon based on the express recognition that some styles of institutional organisation and teaching practice may produce better academic results than others. Whether it be on a theoretical and research level, thanks to contributions from the effective schools movement, or in initiatives of a more practical kind
such as accelerated schools, many authors have insisted on the significance of this school effect and have helped single out some of the variables which explain it (Scheerens, 1992; Sammons, Hillman & Mortimore, 1995).

This emphasis which many authors and movements have placed on the differing contribution which schools make to their pupils' performance, based on different theoretical and empirical analyses, has combined with other current phenomena to place the spotlight increasingly on the internal dynamics of the schools themselves.

The first phenomenon involves the pressure which the current social demand for information and accountability in the public sector is exerting on the education system, one of the most obvious consequences of which is precisely a call for greater transparency. Whilst this demand is addressed to the education system as a whole, it is even more evident where individual schools are concerned, for it is in them that the more general reality of education takes on a specific institutional form. The idea that the black box within which schools are concealed must be opened for all to see is becoming increasingly widespread in today's society (Santos Guerra, 1990).

Secondly, we should recall the central role played by schools in any qualitative plan aimed at improving education. The development of new methods of educational administration and the conviction that passing a law is not enough to make the situation on the ground change immediately have both turned people's attention to schools as the places where any change can and must occur. From this point of view, the notion of a school as a "learning organisation" has once again come to the fore and has been the subject of much interesting analysis (Fullan, 1993).

Finally, another very influential phenomenon must be added in connection with current policies of educational decentralisation (Puelles, 1992): the increased autonomy of the school itself. Although policies of this type vary widely in both conception and implication, they generally agree on the need to give schools greater room for manoeuvre than has traditionally been the case. Nor should we see this as a one-way process, since recentralising endeavours are clearly apparent in countries which already had a high degree of decentralisation. But in all cases, the most widespread tendencies are those which stress school autonomy, whatever their size, in an attempt to enable them to react swiftly and effectively to the demands placed upon them by their particular situation. On the other hand, stepping up their decision-making capacity also requires them to become more accountable for their actions, thus reinforcing the first of the phenomena mentioned above.

A number of different factors have therefore caused the educational researchers, politicians and administrators to focus their attention squarely on
schools. Whereas for many years schools were perceived as no more than the last link in the educational production chain, today they occupy a privileged place in the teaching and learning process and in the strategies aimed at educational change. One of the most visible and interesting consequences of their becoming the focus of so much attention is precisely in the setting up of many different initiatives designed to evaluate them. From being an extravagance promoted by researchers and educationalists, school-based evaluation has now become one of the key components in modern-day education policy, especially where the development of strategies aimed at improving education are concerned.

REGULATING SCHOOL-BASED EVALUATION: FIRST STEPS

The trend towards the development of school-based evaluation which has evolved in many different countries over the last two decades has also reached Spain, although with something of a time lag. In the early eighties, when the first faltering steps were being taken in this direction, Spain was still making a last concerted attempt to achieve full and effective schooling for those who fell within the ages set for compulsory education. As a result, objectives of a qualitative nature were still secondary to the concerns of educational decision-makers. It was not until the end of that decade when, once quantitative requirements were finally being met, that a gradual interest in school-based evaluation emerged.

Nevertheless, it is worth remembering that the 1970 General Education Act had already established the need to evaluate both the implementation of the law and schools themselves. However, despite having been set down on paper, these provisions were never actually implemented in practice, and there was therefore precious little educational evaluation in either the seventies or the eighties.

However, in 1987 the Education and Science Ministry (MEC) published and widely circulated a document entitled Educational Reform Project (A Proposal for Debate) which represented more real progress in this direction. This proposal devoted a whole chapter to evaluation of the educational process, which it considered a key aspect in everyday school life. Although most of the chapter was devoted to the evaluation of pupils and learning, in which the approach adopted was primarily psycho-pedagogical, one of the sections expressly voiced the need to evaluate the teaching processes taking place in schools. The text saw educational evaluation as a task which ran parallel to pupil evaluation and suggested the need for a twofold approach – internal and external. Inspection was to play an important role in this respect, with functions that were to be strictly “evaluative and orientative” (MEC, 1987, p. 149). The document also
advocated the need to conduct “evaluation of the education system, in its
different forms, levels and areas,” which would constitute “true investigation
of its effective operation” (MEC, 1987, p. 150). Although the proposal put
forward some ideas which were new in the Spanish context, it can generally
be classified as relatively unambitious in the area of school-based evaluation,
concentrating more on an analysis of the academic and educational outcomes
than on the organisation and running of schools themselves.

But it was to be in the White Paper on the Reform of the Education System,
published in 1989, where real conceptual progress would be made in the area
of school-based evaluation. Once again, a specific and extensive chapter was
devoted to the subject, in which the principles by which evaluation should be
guided were advocated. In contrast to the 1970 Education Act and the 1987
proposal, which had only referred to the evaluation of educational processes in
a generic manner, the White Paper distinguished between two different levels:
“these two evaluative levels – concerning the education system and educational
processes respectively – must not be confused, although they are clearly related
to one another. Evaluation of how the education system operates is the respon-
sibility and obligation of the education authorities. The evaluation of educational
processes is primarily the responsibility of teachers and students, which are the
players involved in these processes” (MEC, 1989, p. 242).

The document, which was to serve as an educational reform programme, took
a much broader view of evaluation than had been the case in earlier projects.
Whereas these previous attempts had focused above all on the processes of
teaching and learning, on this occasion evaluation was to embrace a far wider
spectrum. Where school-based evaluation was concerned, the White Paper set
out a series of general principles:

(a) Firstly, it underlined the global nature of evaluation. In other words, it was
not to be limited to certain specific aspects of everyday school life, but
was rather to encompass “its projects and achievements, its structures
and resources, the dynamics of its relationships and activities” (MEC, 1989,
p. 243). Without neglecting an evaluation of the students’ outcomes, it
nevertheless moved away from earlier initiatives which had focused solely
on this area, affirming that evaluation “must, as far as possible, be integrated
and should not restrict itself only to the effects or results of educational
processes, but also to the resources with which these were achieved, the
contexts surrounding them, their significance and objectives, and, where
applicable, the structures and dynamic processes which permit the
transformation of material resources and human effort into educational
effects and results which can be evaluated” (MEC, 1989, p. 259).
(b) Secondly, it considered that accountability and qualitative improvement could not be separated: "schools need evaluation of their own operations if they are to be able to regulate themselves properly. Furthermore, families have a right and education authorities a duty to find out how schools are operating" (MEC, 1989, p. 258).

(c) Thirdly, the White Paper maintained the need for and importance of school self-evaluation, although it suggested that this should be combined with external evaluation. The first was to be the responsibility of the school itself, with the school board playing a major role. Its primary objective was to bring added dynamism to the everyday life of the school and guide it towards improvement. The second type of evaluation was first and foremost the task of the inspectorate, which was to redirect its endeavours towards carrying out this type of role.

(d) Fourthly, it proposed adapting evaluation to the particular circumstances of each school. This meant emphasising the use of qualitative methods and implementing the quantitative study in a more complementary fashion.

Overall, it can be said that the White Paper advocated a type of school-based evaluation that was considerably more comprehensive and advanced than had been put forward in earlier documents, thus laying the foundations for the developments that were to take place in Spain in this field in the years directly subsequent to this.

The 1990 Constitutional Act on the General Ruling of the Education System or LOGSE did no more than underline the importance of evaluation, without elaborating on the ideas set out in the White Paper. In its fourth chapter, which was given over entirely to quality in education, it devoted a special article to the evaluation of the education system, considering this to be one of the factors which would contribute to the latter's improvement. It was this article which explicitly mentioned the evaluation of schools and which also created the INCE, to which it attributed a major role in implementing these provisions. In short, whilst the Act cannot be considered to offer anything new that had not already been provided for in the White Paper, it did put its full weight behind the kind of initiative which was beginning to be implemented at the time.

**EARLY INITIATIVES IN EXTERNAL SCHOOL EVALUATION**

It was therefore in the early nineties that school-based evaluation really came into being in Spain, although in some ways it differed significantly with the plans that had previously been made. Whereas the documents drawn up until
that time (particularly the *White Paper*) had placed repeated emphasis on a combination of internal and external evaluation, it cannot be denied that the early attempts that were put into practice on any kind of scale tended to favour this second approach.

The explanation for this phenomenon was that the education authorities considered it necessary to step up external school evaluation in order to bring about evaluation of an internal kind. From this point of view, external evaluation offered a number of interesting possibilities. Firstly, it would contribute to the increased transparency of schools, getting them accustomed to the idea of being accountable for their actions. Secondly, it was to provide them with elements that would help them in creating their own plans for improvement. Thirdly, it was to stimulate the implementation of internal evaluation processes by making clear the advantages of doing this. As a result, the Education and Science Ministry both designed and set up the first initiative of this type with a view to introducing it across the board. The name given to it was the EVA Plan (Luján & Puente, 1996).

Preparation began on this plan – whose innovativeness in the Spanish context must be stressed – at the end of 1990, immediately after the LOGSE had been passed. True, there were valuable precedents, such as that known as the QUAFE-80 model (López & Darder, 1985), which was to become a source of inspiration for many, or the initiative embarked on in 1988 by the Asturias inspectorate (Pérez Collera, 1993). However, none of these had a scope as broad as that achieved by the EVA Plan in the years during which it was applied.

The original project was drawn up jointly by a number of education inspectors and evaluation specialists during the 1990-1991 school year. Its implementation began a year later as a pilot study, focusing on a small number of schools. Although modifications were to be introduced subsequently, the general and specific aims and essential features of the plan remained valid.

Among its characteristics, was its explicitly formative (geared towards school improvement) rather than summative (linked to accountability) nature. Indeed, its overall aim was to contribute to improving the quality of education by improving the organisation and running of schools.

This general aim was given concrete form in a number of specific objectives. As a means of bringing about improvement in participating schools, the plan aimed to set up self-evaluation processes within them, offering them an evaluative model and a set of indicators. The plan also aimed to provide the education authorities with first-hand information on the running of schools, with a view to promoting decision-making. Although the sample of schools evaluated was not a random one, there can nevertheless be little doubt that its diversity and representative nature provided relevant information in helping to draw up an overall picture.
The plan also had a series of complementary objectives. Firstly, it aimed to propagate an evaluation culture in schools, stressing its potential rather than any possible disadvantages. Secondly, it intended it to serve as a testbed for the creation and trial of different approaches and instruments suitable for school-based evaluation, which was particularly important in the Spanish context given its lack of tradition in this field. Thirdly, the plan aimed to help redefine the role of the education inspectorate, assisting in the development of its new evaluative function.

The major features of the plan were also set out in its initial phase. On the one hand, the governing bodies of each school, the teachers, families and pupils would all need to cooperate if it were to work effectively. It was also essential to establish that the objects of evaluation were the organisation and running of the school and not the way in which specific individuals carried out their tasks within it. On the other hand, the plan expressly rejected the classification of schools and emphasised that it was up to each school to reflect on its own situation without comparing itself against the others. Lastly, the final evaluation report was not to be a judgement on the schools but an analysis of its operation, indicating strengths and weaknesses, advocating specific proposals for improvement and promoting undertakings to which both schools and education authorities could commit themselves.

The plan focused on the evaluation of the dimensions and subdimensions listed below, each of which was then subdivided into different variables and indicators (Luján & Puente, 1996, pp. 124–128):

(1) Contextual and personal elements
(2) Projects
   2.1. School's Educational Project
   2.2. Curricular project for particular levels
   2.3. Annual general programme
(3) Organisation and running
   3.1. Governing bodies
   3.2. Teacher co-ordination
   3.3. Teams for specific cycles and teaching departments
   3.4. The school community
(4) Teaching processes
   4.1. Classroom relations and tutorship
   4.2. Methodological aspects
   4.3. Evaluation
(5) Results
Where the evaluation methodology was concerned, the plan adopted an eclectic position, combining quantitative and qualitative methods. Its aim was to gain a closer insight into educational activities in their natural context, without interfering in school dynamics or creating artificial situations. To achieve this objective, a range of techniques and instruments were used, including meetings, interviews, document analysis, observation scales or questionnaires. The participation of a team of three inspectors in each evaluation enabled a three-way comparison of the information collected, thus avoiding any subjective slant. The evaluating team had to compare and contrast their data and analyses with those involved. In order to facilitate evaluation, those in charge of the plan prepared a broad range of instruments (record cards, guides, questionnaires) and a detailed action programme, which some experts found to be overly formalised and lacking in the type of flexibility they deemed necessary in an evaluation process.

The culmination of the evaluation was the drawing up of the final report, and its presentation to and discussion by the school board. Once this exchange had taken place, the final version was approved. It is important to bear in mind that the report was to include proposals for improvement, in the form of specific undertakings which could realistically be taken on board both by the school and the education authorities. This programme of improvement was one of the key components of the plan linking, as it did, evaluation and quality improvement.

In short, between the academic years of 1991–1992 and 1995–1996, the EVA plan was implemented in more than a thousand schools, covering different educational levels. Despite the interesting findings obtained in the course of these five years and the favourable response given to the plan by many schools and by inspectors, it was brought to an end by the change in government in 1996 and was not replaced by any similar initiative.

RENEWED IMPETUS FOR SCHOOL-BASED EVALUATION IN THE LATE NINETIES

However, the experience gathered as a result of the plan had other noteworthy consequences. The most immediate of these was its adaptation by a number of regions (Galicia, Andalusia and the Basque Country), which set up similar plans, although some of these were later abandoned or revised. However, the most important effect was the impetus given to school-based evaluation by the Act on the Participation, Evaluation and Government of Schools or “LOPEG,” which was passed in 1995.

This new law moved a step further on from the LOGSE, by developing some of the aspects which the 1990 Education Act had only touched on. The third
section of the new law was essentially devoted to evaluation and contained provisions covering the different areas of its implementation. It made mention of the Spanish National Institute of Quality and Evaluation, and then went on to deal with the evaluation of schools, teachers, headteachers and the inspectorate. It also covered other aspects such as the professional development of teaching staff in public-sector schools, teacher training and educational research and innovation.

The law emphasised the complementary nature of the internal and external evaluation of schools, in keeping with approaches that had been in use since the 1987 reform project. On the one hand, it established that schools should conduct self-evaluation to assess their own operations at the end of each academic year. On the other, it established that all publicly-funded (whether public or private) schools would be periodically evaluated by the inspectorate.

The Education Act adopted a democratic approach to external evaluation, which required the participation of the whole school community and its governing bodies. In keeping with this principle, the school board had a vital role to play in the evaluation process. Furthermore, external evaluation was to be formative in nature and respect the complexity of the individual school. This meant taking account of the school’s socio-economic context and the resources available to it. It also meant that the analysis undertaken should include not just the results achieved, but also the educational processes. The evaluation criteria and procedures, as well as the general conclusions reached were also to be published to help safeguard the transparency of the process involved.

The “LOPEG” therefore gave solid backing to the external school-based evaluation initiative upon which the Education and Science Ministry had embarked a few years earlier. Subsequently in February 1996, the Ministry was to lay down an order giving more precise guidelines on the implementation of the new school evaluation project which would replace the EVA Plan.

The new regulation proposed an integrated model of internal and external evaluation, which it linked closely together, whilst differentiating between them in conceptual terms. Self-evaluation was to be conducted on an annual basis by the school board and teaching staff, whereas external evaluation would be carried out every four or five years by the inspectorate. One new element was the possibility of including the headteacher from another school on the evaluating team, thus opening the way for peer review. Furthermore, emphasis was placed on the global nature of the evaluation. It was to cover a wide range of issues, including the school’s educational and curricular projects, its annual programming, methods of organisation and operation, educational processes, school atmosphere, complementary and extracurricular activities, the measures taken to cope with pupil diversity and educational results, all put into their
appropriate context. Equally, emphasis was given to the importance of transparency in the external evaluation process. It was therefore established that the different groups in the school community and its governing and coordinating bodies should all participate, that the criteria and procedures to be applied should be published and that the final report should be discussed with the school board in a session specially convened for that purpose.

Overall, the law was a major step forward in systematising the ideas and experiences of evaluation which had been developing in Spain for almost ten years. However, the pre-electoral circumstances in which it was passed were largely responsible for the failure to implement it. The new education authorities did not realise the plan’s real potential and merely discarded it as the work of a previous government. Subsequent action was to revolve more around the implementation of a quality management model for education, based on that of the European Foundation for Quality Management (EFQM). It was this model that had previously inspired certain prominent members of the new ministerial team when drawing up their own proposals (López Rupérez, 1994; MEC, 1997). But the upshot was that the implementation of a school-based evaluation plan comparable to those which are increasingly finding favour in Spain’s neighbouring countries was to be relegated to a later date.

In those regions totally responsible for their own educational affairs however, school-based evaluation was more successful. Pioneers in the field include the Canaries, which will be referred to in more detail in the next section, and Andalusia. More recently, Catalonia can be added to this list, since, at the time of writing, these are the three most interesting school-based evaluation initiatives in Spain.

In 1995, the Andalusian education authorities published and debated the Andalusian Education Evaluation Plan which covered an ambitious set of initiatives. This document dealt with the evaluation of different educational areas, covering everything from the governing and management bodies of the regional administration itself to school support services. Naturally enough, schools were given their rightful and prominent place in this context, and a programme was announced, whose aim was to evaluate them. Although the plan did not go into great detail about how this evaluation should be conducted, it did set out some general principles. These included a clear commitment firstly towards democratic, participatory evaluation and a predominantly formative approach, which did not neglect the need to make the information available to educational decision-makers. It also favoured methodological eclecticism and pragmatism, without espousing a particular paradigm. It adopted a combination of internal and external perspectives and a non-interventionist approach, more inclined to respect the development of school life than to intervene decisively in it.
After a number of pilot studies had been conducted, the Andalusian education authorities drafted and circulated a *Schools Evaluation Plan*, implementation of which began in the academic year 1997–1998. As was the case with the EVA Plan, the aim of the Andalusian project was to introduce an external schools evaluation mechanism as a means to boost the practice of self-evaluation. This line of reasoning is one which we are seeing crop up again and again in the official declarations of the Spanish education authorities.

According to the provisions made by this plan, the inspectorate will evaluate Andalusian schools every four years. The main features of this evaluation are as follows: (a) the evaluation is essentially formative in nature; on a more secondary level, it is also explicative, with the aim of improving the running of the school and increasing its efficacy; (b) it must be multi-dimensional, covering every angle of school life and the way in which these interact with one another; (c) it must avoid comparing and grading schools; (d) it is to be democratic, and must have the participation of all those involved, whilst at the same time taking account of different audiences; (e) it must be part of normal school life, without interfering in everyday activities. It must also be able to provide significant and relevant information to the education authorities and promote the initiation of self-evaluation and improvement programmes in participating schools.

Although the Andalusian plan bears some resemblance to the EVA plan, it is in fact a reworked and improved version, probably created as a result of analysing this scheme. The dimensions evaluated are therefore quite similar to those used in the EVA plan, but are drawn up in terms that are specific to the Andalusian project:

1. Description of schools
2. Planning educational activities
3. Implementing educational activities
   3.1. School government and management
   3.2. Classroom dynamics
4. Results of educational activities

All the indicators selected, evaluation records and instruments used are original. In short, this cannot merely be dubbed an adaptation of the EVA Plan, despite the fact that it was obviously inspired by it and that the two models belong to one and the same family.

Having only just reached the end of its second year of application and with no progress report as yet available, it is difficult to assess the new scheme's results, achievements and shortcomings. Nonetheless, it must be considered an
interesting model, the development of which is worth monitoring closely, since it may lead us down new avenues in the field of evaluation and help extend our knowledge of this area.

The second of the initiatives mentioned was sponsored by the regional authorities of Catalonia. In October 1997, they approved a new regulation setting up an evaluation plan for publicly-funded schools in the region. The model adopted is undeniably original and differs from those previously analysed in a number of areas. Firstly, it proposes a plan to be implemented gradually over a three-year period, in which all the region's publicly-funded schools will take part. Secondly, it combines the goal of institutional improvement with that of accountability, although the latter is targeted more at the education administration than at families or society in general. Thirdly, it revolves around two major evaluation areas which are subdivided into a number of different categories:

(1) Teaching/learning area
   1.1. Results of learning
   1.2. Managing the curriculum
(2) Organisational area
   2.1. Care and monitoring of pupils
   2.2. Managing education activities
   2.3. Functional structure
   2.4. Managing resources

One of the new features introduced by the plan is the conducting of standardised performance tests in and by schools on the basis of which it is possible to draw up indicators of results, without these affecting individual student evaluation. But perhaps the most interesting new element is the implementation of two simultaneous and parallel processes of internal and external evaluation. The plan distinguishes between the two perspectives, attributing the first to the school's governing bodies and the second to the inspectorate. The novelty consists in not simply relying on external evaluation to encourage the internal type, but on both of them developing in parallel, overlapping and complementing one another on a mutual basis. Given the difficulty of the task involved, the intention is for it to be introduced gradually over a period of three years, during each of which a number of internal and external evaluation activities will be carried out. In this way, the decision to undertake an internal evaluation will not merely be left up to schools, but will become an obligation for them.

As is also the case with the Andalusian proposal, the model in question is an attractive one which would appear to be both realistic and feasible. If there
is sufficient political will to carry it out, it should prove an innovative experience and one which will open up new perspectives in school evaluation. The next few years will be decisive in revealing both its achievements and the problems it poses.

NEW APPROACHES TO SCHOOL EVALUATION IN SPAIN

The experiments analysed above have a number of common elements. Firstly, they have arisen on the initiative of the education authorities, who have been the main driving forces behind them. As we have already seen, there are two complementary aims behind this enterprise: firstly to get schools used to the idea of being accountable for their actions and secondly to encourage the introduction of self-evaluation programmes to contribute to decisions which will lead to qualitative improvements. Secondly, they have attempted to combine internal evaluation with external, viewing the two activities as complementary, whilst differentiating between their two respective fields of action. Nevertheless, it is important to emphasise that, in practice at least, more emphasis has been placed on the external than the internal angle. Thirdly, the inspectorate has been given a major role in developing external evaluation. Fourthly, an eclectic methodological approach has been adopted, whereby an attempt has been made to get the most out of each of the methods and procedures used.

In recent times, however, we have seen the development of other types of initiative which have approached school-based evaluation in completely new ways. The earlier models attempted to combine internal and external evaluation, whilst maintaining the distinction between the two and undertaking two parallel lines of action. The novelty which some of these are introducing consists of a new combination of the two perspectives. This type of approach is inspired by certain current models in which external participation is required in internal evaluation, through use of what some authors have called the critical friend or the dialogue for improvement (Nevo, 1997).

The third of the regions mentioned above, the Canary Islands, has opted to develop a model of this type. At an earlier date, in the 1993–1994 academic year, it developed a school evaluation plan similar in style to the ones referred to above, that is, external and using a quantitative methodology (The Evaluation of Schools in the Canaries, 1994). However, this project was discontinued and was soon replaced by another different scheme in which internal evaluation was given the key role. This initiative was promoted by the Canary Institute for Educational Quality and Evaluation (ICEC), which was supported by a number of lecturers at the La Laguna University, who assisted in drawing it up. The
Internal School Evaluation and Development or "EIDEC" Project, was presented in November 1995 and was approved by the Institute's governing board the following March.

The principal new feature of this project was its emphasis on internal evaluation. Unlike the initiatives described above, in this case, the regional authorities focused directly on promoting the internal evaluation of schools, offering their own resources to assist in the task. The ICEC did not see its main responsibility as being to set up external evaluation schemes, but rather to help schools to evaluate themselves. As a result, it invited applications for two successive selection procedures, the aim of which was to choose a group of schools which wished to develop specific internal evaluation plans supported by the education authority. Twelve schools involving different levels were subsequently selected and work began in April, 1997.

In the belief that a number of obstacles to self-evaluation would need to be overcome, the ICEC reached several decisions aimed at assisting in project implementation. First and foremost, an evaluation team was set up in each participating school, on which all the different sectors of the school community were represented. Secondly, it was decided that external experts should join these teams in an advisory capacity. In keeping with this advisory function, it was decided that these specialists should be chosen not from the inspectorate but from school support services. More specifically, it was considered that the most appropriate way to do this would be to use advisors from teachers centres and counselling services. Thirdly, a training scheme was set up for the evaluation teams and external advisors in order to provide them with the conceptual and technical resources required to accomplish their task. Fourthly, a number of instruments were drawn up for the evaluation (questionnaires and other materials), which were then to be made available to schools. The Institute itself would then assist in their quantitative processing and analysis. Fifthly, a project co-ordination group was set up within the ICEC to be in charge of its proper implementation. The internal evaluation process began in April 1997 and has not as yet been completed, its planned duration having been extended until the 1999–2000 academic year.

In line with the general ideas on which the project is based, the true task of evaluation is the responsibility of the evaluation team created by the school itself. The role of the external advisors and the ICEC is more one of support through technical guidance and training. However, it is up to the school bodies to conduct their own analysis, draw appropriate conclusions and make decisions on improvements. The external consultants can have a decisive effect on this task, but they must always abide by the plans set out by the school. Despite the collaboration of external advisors on the evaluation team, this is therefore,
to all intents and purposes, an internal evaluation. It is precisely in this respect that the scheme is a novel and interesting one. As it develops over the next few years, we will have chance to analyse its potential and limitations in greater detail. However, for the time being we can safely say that it opens up new paths for action, at least where Spain is concerned.

Another initiative worth mentioning is a new school-based evaluation scheme which has been in the process of implementation in the Madrid region since the 1996–1997 academic year. Some of its basic features resemble those of the Canary initiative, but overall it is rather different in nature. It has been promoted by a team of teachers and researchers from Madrid universities and some of the region’s secondary schools, including the author of this chapter. The scheme, known as the Secondary School Evaluation Network or “REDES” Project, aims to develop an original model to support the self-evaluation of schools at this educational level. Unlike earlier projects, it has been promoted, not by the education authorities, but by a team of teachers and researchers. Its financial support comes from a private foundation and funds gathered from research projects.

In addition to this fundamental difference with earlier projects, it also differs in other important respects. Firstly, the network concept on which it is based is of particular interest. Indeed, the 34 public- and private schools, which are currently taking part in the initiative on a voluntary basis, form a network which enables the use of results comparison as an evaluation instrument. Although the schools involved are not a random sample of Madrid’s school population, since each is free to decide whether or not it takes part, the selection process has attempted to ensure they are representative so that the network reflects the overall picture of Madrid secondary schools as closely as possible. A comparison of the results obtained by each school against those of the network as a whole gives some idea of its situation in a context such as that of Madrid.

Secondly, participation in the network is based on the principle of confidentiality. This means that the schools taking part do not know which other schools are participating. Furthermore, the only people acquainted with the information obtained on each of them are firstly the members of the external team, who give an express undertaking not to divulge this knowledge, and secondly the school itself, which is free to use the information as it sees fit. This means that, whilst comparison can be used as an evaluation instrument, it is impossible to construct league tables — a practice which many evaluation experts consider to be both harmful and unfair (Goldstein & Spiegelhalter, 1996).

Thirdly, the scheme involves an original compromise between the internal and external perspectives. On the one hand, evaluation is based on the
application of a set of instruments (mathematics, language, natural and social science tests, metacognitive skills and learning strategies, questionnaires on the attitudes, opinions and expectations of pupils, their families and teachers, interviews with management staff, and records of general information on the schools) which are created, processed and analysed by the external team. On the other, it is the schools themselves which analyse their own results, thus consolidating the self-evaluative nature of the process. As a result, although the project is directed from outside, the most important decisions, such as whether to participate in the network and, above all, on the interpretation and use of the data obtained, are the exclusive responsibility of the schools themselves.

Fourthly, the methods used to compile the information are essentially quantitative and standardised, following a model which has already been successfully implemented in other countries (NFER, 1994). Nevertheless, a general model has been developed which takes into account the calculation of the value added by the school. This is done by comparing the results obtained by an individual school against those of the network as whole, but also against those of schools with pupils from a similar socio-cultural background. Data is also collected on socio-economic conditions and on pupil performance on their entry into secondary education, with a view to using these as a weighting element in the final results.

Although the scheme is still relatively limited in scope, since it only began in the 1997–1998 school year, its initial results are nevertheless very encouraging. Firstly, the schools taking part welcome the opportunity to make a non-aggressive, confidential comparison with other similar schools and with a broader context. Generally speaking, this type of information – which is not generally available to schools – is well received. Secondly, the availability of a wealth of detailed information – even if almost exclusively quantitative – has frequently enabled a process of internal reflection which, in some cases, has been supported by a member of the external team. In this respect it would be fair to consider this a true form of self-evaluation. Furthermore, from an organisational standpoint, experience shows that it is possible to embark on a process of self-evaluation with external support, which is neither too time-consuming nor overly burdensome on resources.

Even though this type of initiative combining external and internal evaluation is a relatively recent experiment, whose usefulness is therefore still somewhat inconclusive, there can nevertheless be no doubt that it will help lead the way forward in finding new ways of dealing with school-based evaluation in Spain. More specifically, and from a practical point of view, it will help overcome the separation which frequently treats internal and external evaluation as two opposing, or even incompatible, perspectives. This is a direction it would be
worth taking, in an attempt to open up avenues which are as interesting as they are underexplored.

REFERENCES

EVALUATING ACCELERATED SCHOOLS IN THE UNITED STATES

Pia Lindquist Wong

INTRODUCTION

The Accelerated Schools Project is a national movement of elementary and middle schools committed to the goal of accelerating the achievement of diverse students through the transformation of schools into centers of creativity, critical thinking and powerful learning. Founded over 15 years ago, the Project now extends across 41 states and includes over 700 schools with formal affiliations. The movement is operated in a highly decentralized manner; the National Center is located at the University of Connecticut and twelve regional centers support project development, implementation, and monitoring. In addition, several State Departments of Education have adopted the Project as a process for school transformation, and the Project operates as a pilot effort in Austria, Australia, Hong Kong, and Brazil. This chapter will describe the Accelerated Schools Project and highlight recent efforts to conduct internal assessments of individual schools and third-party evaluations of the movement as a whole.

A PHILOSOPHY AND A PROCESS

The Accelerated Schools Project uses a philosophy of school change to outline a roadmap for transforming school organization and with it teaching and learning conditions, contexts, and outcomes (Hopfenberg et al., 1993; Finnan et al., 1995; Levin, 1991). Research conducted by founder Henry Levin and associates at


187
Stanford University in the early 1980s revealed that schools continued to underserve poor and minority students, despite waves of reforms in previous years. Particular school practices, such as tracking, watered-down curriculum, low expectations, and disconnections between school and home, especially contributed to the poor achievement of underachieving student groups. Key changes in school organization, teaching and learning were identified and integrated to create the three philosophical principles of the Accelerated Schools Project: (1) unity of purpose among all stakeholders in the school community; (2) building on strengths of key actors in the school; and (3) empowering school community members to make important decisions while also holding them accountable for achieving desired results. Undergirding these philosophical elements are core values of equity, school as the center of expertise, trust, communication, collaboration, experimentation, reflection, risk-taking, community spirit, and participation.

The Project’s roadmap for transformation, envisioned as a five year effort, is designed to build a school community’s capacity for institutionalizing the three principles and the core values into the everyday dynamics and practices of the school, from the classroom to the principal’s office. The first year of the transformation process is the most task-intensive. In subsequent years, schools build on the momentum of the launching year as they learn to use the Accelerated Schools process and principles and to deepen their experience using the Inquiry Process and Powerful Learning Framework. The decision to become an Accelerated School marks the first step in the transformation process, and the Project requires that a school engage in a variety of activities to educate themselves about the Project and to determine an appropriate “fit” for their school. Following a buy-in decision of 80 to 90% of the school community, members begin to “take stock” of the current realities of their school. The school as a whole collects and analyzes data as a means of understanding the myriad dimensions of present school life: student achievement, student/teacher relations, student/student relations, budgeting, parent involvement, and the like. Concurrently, the school community reflects on its hopes and dreams for its students. Consequently, teachers, staff, administrators, students, families, and community members use this effort to produce a living vision, not just a vision statement, to guide their school’s transformation. The school community then identifies a limited number of top priorities by comparing taking stock data with the new vision and highlighting the widest gaps between existing realities and desired qualities.

The new priorities constitute the charge of individual cadres, the basic building blocks for the school’s new democratic governance structure. Cadres are heterogeneous groups with representation from all stakeholder groups. Moreover, cadres engage in thoughtful problem-solving based on an iterative
cycle of inquiry and action. In practice, it is the work of cadres following the Inquiry Process that then drives the action agenda of the school – cadres use research, investigation, and observation to delve deeply into school priority areas in order to understand root causes of school challenges (differentiated student achievement, poor student morale, limited parent participation, etc.) and develop studied and effective responses (e.g. curriculum, policies, programs, etc.) that are implemented, evaluated and refined until the desired outcomes are achieved. While the governance structure is being developed, school community members also engage in a theoretical and practical exploration of the Project’s Powerful Learning Framework.

A novice accelerated school will usually have created its new governance structure by the end of the first year; the difficult work of understanding and addressing key priority areas through the Inquiry Process and implementing Powerful Learning in all classrooms comprises the bulk of the work in subsequent years. Typically, the initial work of cadres focuses on understanding the Inquiry Process itself and changing operational paradigms from one of reactive response to one of thoughtful reflection and deliberation. These basic organizational changes often form the foundation on which the more difficult and thorny process of transforming classroom routines, practices and norms occurs, or the implementation of Powerful Learning. The Project’s Powerful Learning Framework has evolved from a set of guidelines that framed instructional planning and curriculum development to a process for creating reflective teaching and learning that emerges out of intensive peer coaching and support teams that operate at the school. Peer support teams have been effective vehicles for focusing the dialogue about student learning on the content, instructional strategies, and context of the teaching/learning experience as well as on five core components of Powerful Learning: authentic, interactive, learner-centered, inclusive and continuous. And through these teams, all members of the school community, and in particular teachers and instructional assistants, have gained a clearer and more unified understanding of instructional expectations, effective strategies, and student strengths. As stakeholders begin to understand this framework, through dialogue, peer observation, and experimentation, they learn to integrate curriculum, instruction, and school organization and context around an enriched approach to learning that accelerates progress towards the school vision and builds upon the three principles.

**ISSUES IN PROJECT ASSESSMENT**

As the number of schools in the Project’s network has grown and as the initial groups of schools have become more sophisticated in their institutionalization
of the Project philosophy, values and governance, questions about the effects of project implementation have emerged at many levels. Participating schools wanted to know “are we doing it ‘right’?” and overseeing districts and state departments (as well as the broader educational community) wanted data about the effects on students and other measures of school improvement. Project leadership at Stanford University focused resources on developing an internal assessment model with a formative focus, and third party evaluators were sought to address summative studies of network schools.

The Project leadership was strongly disinclined to act in an evaluative capacity or to develop an instrument or process that would result in placing schools in competitive positions with each other (e.g. ranking system, accreditation procedure, etc.). A formative assessment process was envisioned as yet another avenue in which accelerated schools could build capacity for the ongoing work of developing a reflective practice and collaboratively transforming themselves. At this formative level, several questions were key. First, to what extent and in which areas of school life have the Project’s philosophy, values and governance structure been institutionalized? Second, how is the internalization of these Project elements affecting decision-making and problem-solving at the school? Third, what outcomes for students and other school stakeholders are resulting from transformed decision-making at the school? Other important considerations in the development of the formative aspect of Project assessment related to maintaining consistency with the decentralized nature of the Project and respecting the efficacy of site based decision-making and governance. That is, Project leadership established two objectives for the development of a formative assessment process: (1) provide feedback and data to schools that would allow them to reorient their efforts vis. a vis. the Project and celebrate successes; and (2) collect data that would identify areas in which Project leadership (at the national and regional levels) needed to provide additional and/or improved guidance, support and direction to network schools.

At the summative level, the issues are quite different. Despite the challenges associated with conducting experimental research on schools (Cave & Kagehiro, 1995; Hollister & Hill, 1995), the Project leadership was in full support of working with third party evaluator to complete an impact evaluation on schools within the ASP network. As soon as this direction was ascertained and an external evaluator selected, the evaluation team and the Project leadership began the task of responding to several central questions including: (1) is there a sufficiently well-defined approach to the Accelerated Schools Project such that schools across different contexts still render consistency in implementation; (2) is there enough of a difference between the ASP approach and that of other educational initiatives to distinguish project schools from other “good” schools;
(3) what is the viability of establishing control groups (e.g. in districts that offer school choice do Accelerated Schools maintain waiting lists that could be used for the purposes of random selection, etc.); (4) is reliable longitudinal data available at the site level (Cave & Kagehiro, 1995)?

During the 1993-1994 academic year planning and development for the internal assessment process was conducted. Staff at the National Center solicited input from all stakeholders in the network. Stakeholders were asked to indicate both what they wanted to know about their school and accelerated schools in general and the ways in which assessment data would be most useful to them. Several state and regional centers had already developed and implemented assessment models and their expertise and experience aided greatly in the creation of the national assessment process. Two working groups were convened of teachers, administrators, scholars and evaluation experts to articulate the salient issues for the assessment process and highlight promising practices and procedures. This comprehensive data collection and reflection period offered several concrete ideas and directions for the formative assessment process.

THE INTERNAL ASSESSMENT TOOLKIT

Building on ideas shared by network stakeholders (Missouri Satellite Center, Louisiana Satellite Center, South Caroline Satellite Center) and incorporating the notion of empowerment evaluation (Fetterman et al., 1995) staff at the ASP National Center developed an “Internal Assessment Toolkit” designed for differentiated use among network members. The Toolkit is intended to be used by schools that have completed the activities of the first year as well as established with some level of stability their governance structure. Typically schools engage with the Toolkit in the second year of project implementation, though some components (e.g. the questionnaire) may be introduced earlier. Moreover, the majority of components in the Toolkit acts as an extension of existing ASP practices rather than the introduction of new elements or tasks. It also mixes qualitative and quantitative data elements. The central function of the Toolkit is to provide schools with a variety of opportunities to assess their progress, reflect deeply about all aspects of implementation, and have those perceptions and reflections cross referenced with structured observations, interviews, data analysis, and artifact analysis. In addition, this process allows satellite centers to more accurately identify ways in which support and mentoring can be given to schools as they implement aspects of the Project.

The Toolkit consists of five data sources: school documents, a coach’s log, a school questionnaire, a school data portfolio and the Accelerated Schools
Benchmarks. Each data source involves a different level of assessment; the toolkit is designed for maximum flexibility such that a school could elect to use only one component or integrate all components together. Each was created to build school community capacity for analysis, reflection, and problem-solving, each of which are important elements in the Inquiry Process and school governance. The components are described below.

**School Documents**
Institutional memory in schools tends to be short and selective. Teachers are so often immersed in the press of responding to immediate student needs (Lortie, 1975) that they have little time for reflection. In addition, a typical teacher's schedule affords few opportunities for dialogue, planning or conversation with colleagues and peers. Because of the pace of a school day, much of the "business" of the organization is traditionally conducted informally, often 'on the fly' as teachers pass in the hallway or chat in the parking lot. This can be effective in some circumstances, but may also lead to miscommunication, misinterpretation, and missed opportunities as good ideas go unrecorded and therefore are not developed and implemented. One of the challenges of becoming an accelerated school is to focus the creativity and energies of a school community on a limited number of priority concerns and to alter the pace so that the efforts of community members are integrated and maintain momentum. At all phases of the first year of implementation (e.g. Taking Stock, Vision, Setting Priorities, Powerful Learning, etc.) schools generate a number of important documents. This continues during subsequent years and takes on even greater significance as school cadres meet to deliberate about challenge areas, gather data in exploration of these challenge areas and develop action plans and evaluations to address them. Thus, schools are encouraged to maintain a record of the activities at the school by carefully organizing important school documents: the school Vision, Taking Stock reports, Minutes from Cadre, Steering Committee and School as a Whole meetings, Cadre Action Plans and Cadre Evaluation results, and other documents (e.g. school improvement plans, newsletters, grant applications). This level of organization can help the school accomplish several important tasks related to assessment: (1) establishes documentation of important institutional events that can be used for reference or reflection; (2) brings schools closer to having collective institutional memory by recording important events and discussions; (3) allows schools to track progress along desired indicators over time; and (4) provides snapshots of the school and its activities at particular points in time.
**School Questionnaire**

The questionnaire was developed by the National Center through a collaborative process that involved review by staff at schools and satellite centers and through a field test at 10 schools in diverse settings and at different points in implementation. The questionnaire is not an evaluation tool but rather a diagnostic device that can provide insight into stakeholders’ perceptions about different aspects of project implementation. The questionnaire addresses several different elements of the Project including: governance procedures, substantive governance issues, school climate and internalization of the Project philosophy and values, familiarity with elements of the Project process, assessment of the operationalization of Project elements (e.g., ‘I participated fully in developing our Vision’), and use of powerful learning frameworks. It is six pages long with a coverpage for demographic information (e.g., grade level, number of years in teaching, etc.) and a final page for open-ended written comments. A five point Likert scale offers respondents opportunities to voice opinions on these different elements and ultimately provides the school community with a snapshot of stakeholder opinion and perspective on the various dimensions of Project implementation and functioning. It is suggested that schools use this questionnaire at least once a year and that results be discussed in a public forum. Ideally, the questionnaire could surface elements of the philosophy or process that a cadre or coach might attend to through a special meeting, workshop or action plan. Future administrations of the questionnaire could reveal the extent to which “problem” areas were being addressed and successes sustained.

**Data Portfolio**

The Project does not promote standardized test scores as an adequate measure of student achievement or knowledge. One concern is that in the case of most accelerated schools, nationally normed rankings may be misleading because of vast differences between their students (typically culturally and linguistically diverse and from poor or working class families) and the norming group (fairly homogeneous in terms of race, class and language) (Campbell, 1999). As a result, the Project prefers using scores to chart the growth of students at a school, that is comparing students to themselves over time (Cave & Kagehiro, 1995) rather than solely comparing them to national norms. Further, schools are advised to assemble a wide range of data that can be analyzed to provide insight into several dimensions of school life. To do this, schools are encouraged to develop a school data portfolio that would contain, at a minimum the following types of data:
(1) information about students including: student enrollment (disaggregated by grade level, gender, ethnicity, language and income level), rate of student attendance, student suspension/expulsion rates, student categorization in special programs (e.g. special education, English as a Second Language, Gifted and Talented Education, etc.), standardized test scores.

(2) information about staff including: staff demographic data by position (e.g. number of African American teaching assistants, number of Latino clerical aides, etc.), staff attendance and turnover rates, and staff involvement in professional development.

(3) information about families and community including: number of family volunteers, family attendance at school events, family participation in such Project activities as cadres or steering committee, and involvement or contributions from other community members (e.g. a business sponsor or partner).

Again, a central goal of the Project as a whole is to create processes and structures through which schools can become centers of reflection, deliberation and informed action. The pressure on educators to distill the essence of a school into a single score is great. Through careful compilation of and reflection on the data in the data portfolio, it is hoped that educators can begin to view their schools comprehensively, as organizations in which many factors – e.g. student attendance, staff development and community involvement – shape that essentialized score.

The National Center requests that schools send copies of their data portfolios to the Center where the data are entered into a centralized data base. Staff and the National Center, where there are often more resources and technological capabilities than at some sites, then compile these data into profiles of school change, along the measures described above. It should be noted that these profiles are not presented as reliable research documents or airtight interpretations of the data, particularly test score data. But analyzing the profiles offers an initial level of data analysis and can give the school rough measures for areas in which changes have occurred – hopefully positive ones.

**Accelerated Schools Benchmarks**

In developing the Benchmarks, the National Center drew on the expertise of several satellite centers, including the Missouri State Department of Education, that had already initiated various forms of assessments, some of which incorporated the use of evaluative rubrics. The Benchmarks are the most intensive form of assessment as they involve comparing a school’s stage of growth with idealized notions of a fully accelerated school. The Benchmarks
are organized around four levels of implementation. Level one is the lowest level of attainment and typically represents a Project component that is underdeveloped. Level four is the highest level of attainment and indicates a Project component that is fully developed and attained consistently and richly. The Benchmarks focus on all of the major Project components: curricular, instructional and organizational aspects of Powerful Learning, the three principles, core values, the establishment and working routines of the governance units (cadres, steering committee, and school as a whole), and the use of the Inquiry Process. The Benchmarks are designed to be used in conjunction with a data collection and analysis process. For each benchmark, pieces of evidence are suggested; it is intended that schools would collect these pieces of evidence, analyze and reflect on them, and use this deliberative process to collectively identify a level of attainment.

For example, one of the Benchmarks addresses "broad participation in school governance" which is an underlying condition for sustaining an accelerated school. To determine how well a school is accomplishing this Benchmark, the school-as-a whole (SAW) would collect and review cadre, steering and SAW meeting minutes for a specified time to consider the scope and breadth of community participation. In addition, the school community members could complete the Questionnaire (mentioned above) where specific items can be cross-referenced as indicators of participation in school governance. Once these data are collected, analyzed and discussed, the school community can come to a consensus decision as to whether they have attained a level 4 (a clear majority of school community members consistently participate in school governance through membership on cadres and/or steering committee), a level 1 (few teachers and staff consistently participate in school governance) or somewhere in between (level 2 or 3).

Schools are advised to use the Benchmarks with caution. Because of the large number of accelerated schools and the diverse contexts in which they operate (rural, urban, suburban, etc.), it is nearly impossible to develop a rubric that will be wholly relevant to all schools in the movement. In the same vein, it was with some trepidation that staff at the National Center endeavored to "define" a top-performing accelerated school, primarily because it is understood that each school will implement the Project in its own way and that some schools will excel in certain areas, and may even surpass by a large margin the expectations of the National Center for a particular Project component.

*Coach's Log*

Coaches or trainers play a pivotal role in Project implementation and support. The configuration of coach and school varies from region to region, however,
most schools do work with a coach/trainer on a weekly basis at least for the first year of implementation. The coach is responsible for planning and conducting all of the major staff development activities related to Project implementation during the first year. The coach must keep the process developing smoothly while also paying close attention to the changing school community dynamics, the ways in which individuals are understanding and internalizing Project values and principles, and the extent to which Project components are understood and operationalized by different individuals. The coach must often play the role of teacher, mentor, and support provider simultaneously, and occasionally, the coach ‘runs interference’ between principal and staff, staff and families, and school and district! While the coach becomes fully immersed in the transformation process of his/her school, s/he also must maintain a posture of objectivity in order to determine the nature, scope and content of the different professional development activities to be provided. And, coaches also need input and feedback on the progress that is being made at their schools.

Because of the challenges associated with being an effective accelerated school coach, the final tool in the Toolkit is the Coach’s Log, which is an organizer that helps coaches reflect on their own work and the progress of their school as well as develop strategic planning skills for effective coaching of their school. The coach’s log contains various ‘worksheets’ that frame issues for reflection and analysis. These worksheets loosely follow the phases of Project implementation so that the coach can use them to develop a deeper and more comprehensive understanding of their school as the relationship unfolds. The worksheets include such topics as: (1) a school personality test in which the coach makes inquiries about factors that shape school culture (e.g. teacher turnover, new feeder schools, new curriculum), staff demographics and experience, staff morale and relations, staff, principal and school community strengths and challenges; (2) summary of the Taking Stock process including the major strengths and challenges identified and the coach’s assessment of the ways in which the school community was able to work cooperatively and efficiently; (3) a similar summary of the Vision process focusing on the ways in which all stakeholders participated and the benefits or conflicts that may have emerged through the process; (4) observations of the different levels of school governance including challenges that any might be facing in meeting management or in understanding the inquiry process as well as strengths of each level; and (5) coach’s observations in classrooms and identification of powerful learning elements that are emerging.

It is expected that the coach’s log will be used in a collaborative way rather than by individual coaches solely on their own. The National Center and
affiliated satellite centers organize annual retreats that provide professional development and networking opportunities for coaches. Elements of the log can be used as reflection, discussion and activity tools at these events as well as operate as points of dialogue for coaches and their mentors. In addition, portions of the coach’s log can be utilized to spark discussion with the coach’s school and elicit feedback from community members.

APPLICATIONS OF THE TOOLKIT

An underlying assumption of the Toolkit was that it would be adopted as a totality or by component depending on the needs and objectives of particular schools and centers. In addition, it was expected that elements of the Toolkit might be modified to best suit the objectives of particular sites. Thus, full implementation of or complete fidelity with the Toolkit was not a central goal. The main purpose was to provide a set of guidelines and tools for evaluation that could serve as a resource to schools. The only effort to centralize use of the Toolkit has been a general call for School Data Portfolios so that a database of participating school demographics could be established – with networking and access potential for all Project members (National Center, 1996–1997). Practice has upheld these initial assumptions, and reports from various sites confirm that the use of the Toolkit has been shaped by local needs and interests. For example, the Ohio Satellite Center, which works with schools in Ohio, Maryland, and Indiana only uses the Benchmarks. In this context, the Benchmarks are used in two ways. First, they are a reflection tool, primarily among principals, at a Principals’ Institute and with coaches, during their training and retreats. Second, a version of the Benchmarks is used to complete an evaluation of schools in their second year of implementation. Interestingly, this satellite center has also integrated elements of the School Questionnaire and Data Portfolio with state requirements for school reports, and provided this form online for accelerated schools.

OTHER EVALUATION EFFORTS

Many participating schools and centers had begun to make inroads into developing assessment tools for the Accelerated Schools process and philosophy prior to the completion of the Toolkit. As always, the practice at the grassroots level outpaced centralized efforts in terms of initiative and ability to integrate with current mandates and requirements. Schools in Iowa and Florida developed
innovative assessments that included rubrics and questionnaires (Newsletter, 1995–1996). Satellite centers in South Carolina, Colorado and Massachusetts created such assessment tools as: (1) infusing state mandates for learner standards with rich descriptions of powerful learning and other elements of the Project philosophy and values and overlaying Project processes such as Inquiry and the work of cadres onto state requirements for planning and goal setting; and (2) supporting schools as they developed action research projects to investigate school changes and improvements related to Project implementation (Newsletter, 1995–1996).

Plans for third party evaluation continue to be developed though the challenges to designing a rigorous and experimental evaluation have proved daunting. Manpower Development Resource Corporation (MDRC) approached the Accelerated Schools Project in the early 1990s as a potential third party evaluator. At this time, they conducted a thorough analysis of the potential for completing an experimental research effort examining the impact of the Project on a range of indicators related to student growth and achievement, foremost, and teacher change and growth secondarily. This preliminary research revealed the Project and its affiliated schools to be appropriate prospects for such an evaluation and it has proceeded forward. Preliminary findings, presented at the 2001 Annual Meeting of the American Educational Research Association (cf. Doolittle, 2001; and by Bloom, Ham, Melton & O'Brien, 2001), indicate that schools in the study experiencing full implementation of the Accelerated Schools Project (that is, in their fourth and fifth years) can attribute improved student achievement to Project implementation.

SUMMARY

Members of the Project's vast network are excited and challenged by the task of assessing their work in accelerated schools. Though many have experienced assessment as yet another set of demands, most have realized the value of reflection and analysis and are energized by the revelations and insights that assessment can provide. The stakeholders in Accelerated Schools, students, teachers, administrators and families, all share deep concerns about and commitment to the creation of successful, enriched, and vibrant schools. Assessment strategies, as embodied in the Accelerated Schools Toolkit and other efforts detailed above, provide important avenues for identifying information and data necessary to energize the transformation of school communities and sustain their stakeholders' commitment.
Evaluating Accelerated Schools in the United States

NOTE

1. For the most current information on the Accelerated Schools Project, please consult: www.spuconn.edu/~wwwasp/main.htm

ACKNOWLEDGMENTS

The author would like to acknowledge the valued contributions of ideas, materials, and strategies from: Tselane Sheppard-Williams, Stuart Yeh, Simeon Slovacek, staff at the National Center, and staff at the Missouri, Louisiana, and South Carolina Satellite Centers, and staff at Aurora Public Schools in Colorado. My appreciation to Henry Levin and Gene Chasin for their helpful editing suggestions although any flaws in the manuscript are mine alone.

REFERENCES

SCHOOL-BASED EVALUATION AND THE NEO-LIBERAL AGENDA IN CANADA

Les McLean

INTRODUCTION

I set out to write a chapter like the others in this book, describing and discussing examples of school-based evaluation in Canadian schools. “Delighted,” I told the Editor; “With the networks built up in over thirty years doing research and evaluation in all ten provinces and one of the three territories, I will find enough cases to yield a good story.” Again and again people replied that yes, they had begun programs that would fit the description of school-based evaluation (SBE), though not often given that exact label, but that budget cuts, more and more external testing and repeated curriculum changes in the name of reform meant that all efforts at grass-roots evaluation had been abandoned. As my search continued, I became increasingly desperate, looking into every activity with a name that began with “school-based.” There are many school-based programs for health promotion (reduce smoking, . . . ), violence prevention, outdoor safety and the like, but these rarely involve school personnel in design or delivery or evaluation. School-based management (Brown, 1990) is often mentioned but turns out to be just that: budget setting and administration. Finally I found a large program whose objectives suggested that SBE might be a part. It is not, but that itself fits the story I have to tell – how neo-liberalism is degrading public education in Canada. First, a word about my understanding of SBE.
THE DEMOCRATIC NATURE OF SCHOOL-BASED EVALUATION

In a review (McLean, 1997) of School-based Evaluation (Nevo, 1995), I characterised the process as:

(1) communication, motivation and skill-building through in-service training of individuals removed from schools for the purpose; (2) spreading the word and the spirit by means of evaluation teams inside receptive schools, teams that take up tasks supported by (perhaps suggested by) evaluation specialists – learning by doing; (3) planting the process firmly by naming permanent teams and ensuring rotating membership, including one or two advisors; and (4) including external evaluation openly and explicitly, perhaps bringing it into the process of implantation (p. 114).

Steps 1–3 are just not possible under current financial and administrative constraints. In-service training has been cut back almost to nothing, and what remains must be used for minimal familiarisation with new report cards, curricula and large-scale tests. Teaching loads have been increased by one period per day in Ontario, leaving insufficient time for lesson preparation – and no time at all for work on “evaluation teams.” There are no funds for “advisors,” and teacher turnover is increasing, accelerated by early retirement incentives. Attention to external evaluation is strongly encouraged, but there is little or no guidance how this might be done. “The school system redesigned for global competitiveness is accountable for aggregate test scores, rather than for ensuring that each student can develop to his/her potential” (Kuehn, 1996). Sherman Dorn has argued eloquently that external testing and control have left a “political legacy” in which “testing defines legitimate discussion about school politics” (Dorn, 1998). Instead of a school-based democratic process, the present climate demands a top-down administration of decisions made at the level of the provincial Ministry of Education.

An ironic exception to the top-down model is the encouragement given to partnerships between schools and the private sector. This can take place at the level of the local authority (such as granting exclusive rights to one soft-drink company, an agreement that recently gained a large local authority in Ontario a payment of C$20 million), but individual schools can also make deals. For example, dozens of secondary schools accepted installation in their school of a new television network linked to a satellite dish and a multi-media computer laboratory (YNN, 1999). Schools received the expensive hardware at no cost in exchange for an agreement to show a 12.5 minute news and current events program in all classrooms each day, a program containing 2.5 minutes of “sponsored messages” (e.g. health advocacy and commercials sponsored by Nike and the like).2 United Statesians will recognise this as the Canadian version
of Channel One. Schools in six provinces have signed up, in spite of official opposition from three provincial governments and all of the teachers’ unions.\(^3\)

Let me turn to the one program I thought might embody SBE.

**THE MANITOBA SCHOOL IMPROVEMENT PROGRAM**

The Manitoba School Improvement Program (MSIP) was established in 1991 by (surprise!) a Canadian charitable foundation, the Walter and Duncan Gordon Foundation. The Foundation chose the Canadian province of Manitoba as the pilot site for a Canadian secondary school improvement project because: (a) they were interested in enhancing education for “students at risk;” (b) they were looking for a province with an educational community that would welcome and support their involvement; and (c) they wanted to start in a location that was manageable within the constraints of their desired budget allocation. Manitoba met these criteria and was willing to take part in the innovative programme. Established initially with an urban focus, the program has expanded to include some of the province’s rural and northern schools. The Conservative government of the day did not have to put up a penny, but it may be relevant that the New Democratic Party government that succeeded them has agreed to make the MSIP a “self-sustaining educational partnership” (Earl & Lee, 1998, footnote 4).

Each school had to prepare a pre-application plan that included developmental aims, objectives and resource implications, as well as budget and evaluation designs. The activities were required to be project-based. These plans were undertaken in conjunction with an MSIP co-ordinator and evaluator and each MSIP school was assessed on a common set of criteria (Earl & Lee, 1998). The spirit of SBE appears when we learn that projects were required to:

1. be *school based and teacher initiated* (emphasis added),
2. focus on the needs of the adolescent students,
3. address fundamental issues of educational improvement and *student learning* for at risk students, (emphasis added)
4. have the potential for long term impact on the school,
5. be designed or developed to *incorporate a collaborative and participatory approach within the school* (emphasis added), and
6. *include an appropriate evaluation component* (emphasis added).

The generous foundation grant provided funds for teacher released time as well as the co-ordinator, an external evaluation team and an Educational Advisory Committee. Comparison with the four aspects of SBE stated above suggests the potential for strong overlap, but a careful reading of the external evaluation
report (Earl & Lee, 1998) reveals this not to be the case. School marks and test scores were not published and in several cases were not available to the external evaluators. School reports offered opinions whether achievement had improved with the advent of the MSIP. In spite of objective 5, there was only modest involvement of teachers in the projects and collaboration among them. Involvement was rated "High" (on the basis of a teacher questionnaire) in only 3 of 13 "senior" schools and in only 2 of 7 "mid-years" schools. Collaboration ratings were slightly lower (Earl & Lee, 1998, Appendix IV). In Canada, at least, it appears that even substantial external funding does not lead to school-based evaluation. Nevo's book did not appear until 1995, but articles were in journals before then. Disappointed, we return to the real educational worlds without foundation support.

**EDUCATIONAL CONTEXTS IN CANADA SINCE 1990**

Nova Scotia, New Brunswick and Prince Edward Island recently elected Conservative governments, replacing long-standing Liberal regimes, and "Liberal" parties in the other provinces are increasingly more conservative than the Conservatives. The pattern is evident in all provinces, but in none more than in Alberta, British Columbia and in the one I inhabit and know best, Ontario. Here are a few of the changes made by the Ontario Conservative government, recently re-elected to a second term with another strong majority.⁴

- Regulation of the elementary and secondary teaching profession was given to the "College of Teachers," whose governing board is appointed by the Minister of Education. Entry to teaching still requires a certificate issued by university Faculties of Education, but the College of Teachers (mandatory membership) is responsible for continued certification and for discipline previously handled by elected school boards (local authorities).
- The number of local authorities has been drastically reduced by forced amalgamation and their powers reduced by removal of the right to raise funds through taxes, centralisation of authority in the Ministry of Education and devolution of decisions to the school level. Local authorities still have heavy responsibilities but little authority, and the provincial education budget has been cut every year even as enrolment grows. (New Brunswick eliminated local authorities completely and replaced them by parent committees and advisory councils responsible to the district superintendents and ministry – not to the public electorate.)
• Every school must have a “School Council,” made up of parents, teacher representatives, officials and community members, but ultimate power and authority lie with the Minister.
• Principals and Vice-Principals have been removed from the Federations (the teachers’ unions) and forbidden to organise.

Perhaps a vignette will illustrate the sort of climate that has emerged in Ontario schools:

1. The Superintendent (a highly placed official) of a local authority posts a note in a computer-based “chat space,” set up for communication among teachers and officials. The note says that supplies of paper and materials are short for students because the teachers have been too greedy and taken them for themselves.
2. The union representative in the school (every school has one) posts a polite and respectful rebuttal, arguing that supplies are short because the budget has been cut to the bone and that teachers have very few resources for themselves.
3. The union representative discovers that she can no longer post notes in the chat space – no notice, no warning.
4. A teacher posts just this note: “Who was it said, ‘I may disagree with what you say, but I will defend to the death your right to say it’?”
5. The teacher discovers that she may no longer post notes – no notice, no warning.

My argument is that such a context, commonplace rather than unusual, is antithetical to the essentially democratic process of school-based evaluation, and that the context has emerged, in Canada at least, because society is under the influence of an economic philosophy with the oxymoronic name of “neo-liberalism.” Let me describe what I understand neo-liberalism to be.

THE NEO-LIBERAL AGENDA

Neo-liberalism is a set of economic policies that have become widespread during the last 25 years or so. Although the word is used less often in North America than in Europe, you can clearly see the effects of neo-liberalism here as the rich grow richer and the poor grow poorer.5

“Liberalism” can refer to political, economic, or even religious ideas. In the USA. and Canada, political liberalism has been a strategy to prevent social conflict. It is presented to poor and working people as progressive compared
to conservative or rightwing. Economic liberalism is different. Conservative politicians who say they hate "liberals"—meaning the political type—espouse enthusiastically economic liberalism, including neo-liberalism.

A memorable definition of neo-liberalism came from Subcomandante Marcos at the Zapatista-sponsored "Inter-continental Encounter for Humanity and Against Neo-liberalism" of August 1996 in Chiapas when he said: "what the Right offers is to turn the world into one big mall where they can buy Indians here, women there..." and, he might have added, children, immigrants, workers or even a whole country like Mexico (Marcos, 1997). The main points of neo-liberalism are:

- **The rule of the market.** "Free" enterprise or private enterprise must be liberated from any bonds imposed by government (the state) no matter how much social damage this causes; greater openness to international trade and investment (as in the North American Free Trade Agreement, NAFTA); and strategies to increase profits by de-unionising workers, eliminating workers' rights that have been won over many years of struggle and most dramatic of all, eliminating thousands of jobs entirely. In education, partisans argue for "school choice," often in the form of "vouchers" (money or paper given to parents, guardians or students of legal age with which they can buy a place in the school of their choosing) and "Charter schools" (independent, but publicly-funded, schools organised around some special group or interest).

- **Removing price and currency controls.** There must be total freedom of movement for capital, goods and services, especially the first. The rationale offered for this is, "an unregulated market is the best way to increase economic growth, which will ultimately benefit everyone." Examples include Reagan's "supply-side" and "trickle-down" economics; some argue that there is now some "trickle-down" happening in the USA, more than a decade later, but the gap between rich and poor has steadily grown and is still growing.

- **Cutting public expenditure for social services.** These include education and health care, reducing the safety net for the poor and even maintenance of roads, bridges, and water supply—again in the name of reducing government's role no matter what the consequences. In a burst of honesty, the United Statesian political figure and sometime Republican presidential candidate, Jack Kemp, called it "root canal economics." A very few neo-liberal writers, such as Canadian-born David Frum (Frum, 1994), oppose all government subsidies, including tax benefits for businesses, but the majority on the right have ensured that those subsidies remain or are increased.

- **Deregulation.** In practice, this means reducing government regulation of everything that could diminish short-term profits, including protecting the
environment and safety on the job (even though these can be argued to enhance profits in the long-term). It also includes encouragement of private universities, as well as secondary and elementary schools where management is free to employ non-certified teachers (usually under the rubric of Charter Schools).

- **Privatisation.** State-owned enterprises, goods and services are sold or contracted-out to private investors. Included are banks, key industries, railroads, toll highways, electricity, schools, hospitals and even fresh water. Although always done in the name of greater efficiency, which is often needed, privatisation has mainly had the effect of concentrating wealth even more in a few hands and (as in the privatisation of British Rail) making the public pay even more for reduced services and less safe operations. Enormous enterprises, in which the public has invested billions over many years, are sold for a few hundred millions.

- **Eliminating the concept of the “public good.”** This concept has been replaced with “individual responsibility.” The poorest people in a society are expected to find solutions to their lack of health care, education and social security along with the middle-class and well to do. The concept has been argued for education as well (Frum, 1996), putting responsibility primarily on “the family” rather than provision of free education open to all.

Around the world, neo-liberalism has been strongly encouraged by powerful financial institutions such as the International Monetary Fund (IMF), the World Bank and the Inter-American Development Bank, whose common denominator is their lack of transparency and democratic accountability. It is the dominant political and economic philosophy over the Americas, North, Central and South, sometimes called “structural adjustment” (George, 1999).

An early and dramatic example of neo-liberalism at work came in Chile (thanks in no small part to the theories of University of Chicago economist Milton Friedman), after the coup against the popularly elected Allende regime in 1973. The coup has since been shown to have been aided and encouraged, if not initiated, by the U.S. Central Intelligence Agency. A small chink in the armour appeared more than 25 years later when the leader of the coup, General Augusto Pinochet, was detained in England awaiting extradition to Spain to be tried for crimes during his regime. The initial recommendation by the House of Lords not to send Pinochet back to Chile was heartening in itself, but he was sent back and few believe he will ever be tried.

Effects soon appeared in other countries, some of the worst in Mexico, where wages declined 40 to 50% in the first year of NAFTA while the cost of living rose by 80%. Over 20,000 small and medium businesses failed and more than
1,000 state-owned enterprises have been privatised in Mexico. The crowded and unhealthy assembly plants that have sprouted along the U.S. border have not begun to make up for the losses. As one scholar said, “Neo-liberalism means the neo-colonisation of Latin America.” It first made its appearance in a pure form in New Zealand in the 1980s, under the leadership of the (now infamous) Minister of Finance, Roger Douglas (Clancy, 1999). A new government in New Zealand is trying to undo some of the damage, but they face a daunting task.

In the United States and Canada, neo-liberalism has taken over national, provincial and state governments, whose legislators feel they must cut welfare programs, attack the rights of labour (including all immigrant workers), and above all, reduce the costs of health care and education. It came as a great shock to most of us when the political party seen as most sympathetic to social services and labour (the New Democratic Party) unexpectedly came to power in Ontario in 1990; they set about budget cutting and adopted other neo-liberal policies as draconian as any Conservative. Some critics go so far as to argue that “the neo-liberal agenda threatens the existence of Canada” (Shields & McBride, 1997). When adopted piecemeal, as in the Canadian provinces, fractures and discontinuities of structural adjustments appear, but the local versions of neo-liberalism are no less real as a political phenomenon and no less consequential to the lives of Canadians than the unitary version was to the people of New Zealand (Clancy, 1999). In the USA., the Republican “Contract on America” was pure neo-liberalism. Its supporters worked hard to deny protection to children, youth, women, the planet itself, in the name of cutting taxes and reducing the size of governments. The beneficiaries of neo-liberalism are a very small minority of the world’s people. The vast majority is poorer, less healthy and has less access to quality education, and yet the advocates “have made neo-liberalism seem as if it were the natural and normal condition of humankind” (George, 1999).

It is ironic that in principle, neo-liberalism and school-based evaluation are more in agreement than in conflict. Local initiative and control are valued aspects in each perspective. Accountability measures that include external testing are explicitly endorsed by both. Parent involvement is highly prized, if not essential. Central control over the curriculum is by no means necessary for school-based evaluation, but neither is it incompatible with it. The two diverge in practice, however, in the unwillingness of neo-liberals to cede real power and decision-making to local committees. The budget restrictions integral to the neo-liberal agenda are designed as much to curb local initiatives (except business partnerships) as they are to save money. Expenditures reduced by cutting some categories of staff are offset by extensive (and expensive) testing programs, and as many costs as possible are transferred to the parents or made
the responsibility of school fund-raising. The net effect is as described before: teachers and officials cannot take the time required to organise and implement school-based evaluation. In the spirit of this book, I describe how a Canadian neo-liberal school-based evaluation would almost certainly turn out.

**HYPOTHETICAL CASE**
**(A COLLECTION OF ACTUAL EVENTS)**

Mackenzie King Secondary School has about 1700 students in grades 7–12 coming from a variety of ethnic and cultural backgrounds. About a quarter of the students come from homes where an Asian language is spoken, but by and large they have been in Canada for five years or more and have good language proficiency. Located as it is in a mixed middle- and upper-middle suburban community, it has no more than the average number of behaviour problems and special-needs students, and it has long had a "gifted program" for students who score more than one standard deviation above the mean of a common published IQ test. Just over half of the graduates go on to post-secondary education, either to college or to university. The new School Council is beginning to function under the leadership of Principal Lorena Medwall, who has excellent relations with the teachers, the students and most of the parents. About a dozen parents are annoyed with the Principal and the local authority because the school has accepted a television network and multi-media computer laboratory in exchange for showing a 12.5-minute television program each day. The program contains 2.5 minutes of public health and safety messages and/or commercials. The school and the local authority are conducting their own evaluation of the presence of the television programs and computer lab in the school, and an external evaluator is preparing a report on the private sector initiative for five schools across Canada, but the opponents of the initiative are adamant that no benefits can possibly compensate for the intrusion of commercial messages in the school. The issue remains active and some of Principal Medwall’s time is taken each day responding to objections. Apart from this issue, the school is operating with only the usual chaos.

Three years ago, the Minister of Education announced new external examinations, beginning with literacy and numeracy in Grade 9, and the local authority urged schools to step up their own efforts to show accountability to the public for good performance. The external examinations are developed and administered by an "arm's length" agency, the Education Quality and Accountability Office (EQAO), set up by the Ministry several years ago, but EQAO involves teachers in all stages of development, from item/task writing to the marking of performance tasks. Extra work thus devolves onto two or
three teachers in each school. At the same time, the Minister decreed that the teaching load would be increased. Teachers in some local Authorities went on strike (citing the extra load and other issues) and most began a program of “work to rule,” but these efforts were not successful in reversing the Minister’s “reforms.” It was in this atmosphere that the school had to decide what, if anything, they could do to show greater accountability. The Head of the Mathematics Department, who was enrolled in a Master’s degree program, suggested that they consider a process he had studied: School-based Evaluation (SBE). He promised a written description for the next Head’s meeting.

When the Heads learned what SBE entailed, their reactions ranged from depression to rage. How could they possibly undertake such an ambitious effort over several years under current conditions? Regional in-service training? Internal evaluation teams – teams that become a permanent part of the structure? Convincing the EQAO to postpone provincial testing? The Principal agreed that SBE was impractical to undertake in full but reminded them that it was certainly undesirable to do nothing and suggested that they see if there was any version of SBE that would: (a) be feasible in their school; (b) might satisfy the officials in the office of the local authority; and (c) would be endorsed by the parents (including the dissidents currently at war over the commercials). The Mathematics Head and two others agreed to prepare a proposal to put to all the teachers after circulation to Heads and approval by them. They promised that the proposal would contain several compromises and if accepted by a strong majority of teachers would be taken to the School Council for approval.

• Inservice training would not require any teachers to be removed from the school during school hours. If possible, a small group would be recruited to attend one or two evening or weekend sessions with the instructor in the Math Head’s graduate course. This group would form the core “evaluation team” that would lead the SBE.

• The Mathematics Head would serve as the evaluation leader, advised by the instructor at the university. (The Principal was confident that a sufficient honorarium could be obtained from the office of the Local authority to ensure participation by the instructor.)

• Members of the core evaluation team would recruit teachers to form at least two or three continuing evaluation teams for different areas of the school curriculum. Students would be involved as appropriate and feasible. No promises would be sought of continuing membership on evaluation teams.

• The Principal would seek advice from a measurement specialist, perhaps from the Research Division of the local authority on integration (implantation?) of the EQAO tests in the schools SBE plan.
The high regard in which the Mathematics Head and the Principal were held was confirmed when the proposal was approved at a full teachers’ meeting with no negative votes and only four abstentions. Approval was made conditional upon: (a) acceptance by officials of the local authority and provision of the required honorarium; (b) approval by the School Council with no negative votes, and (c) review of the whole process each year and cancellation on a vote of a simple majority of teachers. The School Council was enthusiastic and approved the proposal unanimously, the dissidents speaking strongly in favour. (It became clear that they believed that such an evaluation would reveal harmful effects of the mandatory television programs and commercials). It was decided to proceed with planning and organisation even before hearing from the local authority, since it had become more and more difficult to get responses from them about anything.

Stage 1

The instructor in the Math Head’s graduate course (a keen advocate of SBE) agreed to lead at least four sessions at the school for the Principal, the Math Head and teachers who volunteered to participate – whether an honorarium was forthcoming or not. The SBE “formula” suggests 70–80 hours of training in program evaluation, testing, data collection procedures and data analysis (Nevo, 1995, p. 50), but the group agreed that the equivalent of four half-days (16–20 hours) was the maximum they could undertake. They decided to concentrate on program evaluation and data collection procedures, hoping that the measurement specialist would help with the rest. At the urging of the Principal, who promised unspecified released time at some time in the future, six teachers volunteered and professional development sessions were scheduled: two three-hour evening sessions and two Saturdays, so that at least 18 hours could be planned. This was quite a vote of confidence by the teachers, who had been on strike in the fall and working-to-rule until very recently. A key element was the exclusively local nature of the initiative, given the distrust existing between teachers and local authority and outright hostility to the provincial authorities.

The instructor presented them with 9 “lessons in SBE,” which were discussed during the first evening session and all of the first Saturday (Nevo, 1995, pp. 51–59):

(1) We understand best the meaning of evaluation through the distinction between description and judgement. Evaluation was defined as a systematic activity that uses information to describe educational objects and judge their merit or worth. (The goal to judge merit or worth provoked spirited debate, but it was agreed in the end to accept it subject to later consideration.)
(2) Students and their academic achievements should not be the only aspects considered in school evaluation. (Enthusiastic agreement with this ‘lesson’)

(3) Looking only at outcomes or impacts (e.g. ‘outcomes-based evaluation’) was not at all sufficient for school evaluation. (‘Hurrah,’ the Principal said.)

(4) On the other hand, SBE has to yield information both for planning and improvement and also for accountability outside the school. (This was immediately agreed in principle, but about two hours were spent in discussion of ways to achieve it. Finally, it was agreed that this had to be an ongoing discussion, with assistance from the measurement specialist.)

(5) No single criterion, or even one set applicable to all schools, could be found to judge the overall quality of a school (or a teacher or a program). More important, there is no need to find one. (‘Hurrah!’ the teachers said. ‘Finally someone out there has said this.’)

(6) Evaluation of most use to the teachers and students in a school is best done internally, supported by planned professional development involving technical assistance from outside where appropriate. (The teachers commented that they had already learned this lesson, or they would not be at this session.)

(7) No school staff can carry out a meaningful evaluation unless they assemble a variety of tools and methods from various sources and adapt them to their own needs and capabilities. (The first response to this ‘lesson’ was a thoughtful, if not stunned, silence. The instructor, from long experience, said nothing for a few minutes and then said, ‘Let’s look at a few examples.’ Discussion continued into the next session, recognising again that this was a topic ‘to be continued.’)

(8) Learning-by-doing is still the best way to learn how to do evaluation. (‘We’re convinced!’ everyone said.)

(9) Evaluation done in the school, involving the whole school community, must be completed before any external evaluation (e.g. provincial testing) can be meaningful and useful. (Everyone agreed that internal evaluation was essential. Most were sceptical that this could lend meaning to provincial testing, but they agreed to try to keep an open mind.)

The instructor then described the tasks to be undertaken in Stage Two: Organisation of evaluation teams and selection of an “evaluand,” a project, a program or some other aspect of the school’s mission that can be identified and evaluated in a reasonably short time. The purpose of this evaluation is to practice and refine their evaluation skills.
Stage Two

Two teams were formed: one to evaluate the cafeteria supervision scheme (recently implemented) and a second to evaluate implementation of the Grade 9 unit in Probability, introduced last year as required by the provincial curriculum guidelines. One academic and one non-academic evaluand were chosen to bring out the different requirements of each – if any. The term “evaluand” was suggested by one of the teachers, Sheila, a mathematics teacher who was at the moment attending a graduate evaluation course (though she had told only a few close friends about it). She proved to be a prolific source of ideas and concepts, just in the process of being gleaned from the course and the textbook, Program Evaluation: Forms and Approaches.10 Sheila enriched the evaluation tasks by suggesting that a “Clarificative” form was most appropriate for the cafeteria supervision evaluation, while a combination of Clarificative and Interactive forms would be useful in the evaluation of the Probability unit. At the heart of the Clarificative form is description of the theory or logic model underlying the evaluand, a concept that intrigued two of the teachers enough that they agreed to undertake the description for the cafeteria supervision scheme. The Mathematics Head agreed to provide a draft theory/model for the Probability unit. Step 7 from Stage One was unexpectedly underway.

By the end of that school year, the teams had produced rough reports on cafeteria supervision and the probability unit and were ready to move on to Stage Three: Evaluation of significant aspects of the school program. They had not drawn on outside technical assistance (other than the instructor for the professional development sessions), finding resources within the school that they had not known about when embarking on their version of SBE. When they parted for the two-month summer break, they had agreed to undertake evaluation of two programs during the next school year: the Gifted Program and the Grade 10 “transition” curriculum in mathematics (another new program required by the recent provincial curriculum ‘reforms’). Over the summer, the Principal would attempt a logic model of the Gifted Program, and Sheila (who had now completed her evaluation course) would attempt a description of the theory underlying the Grade 10 transition program as a project in one of her two summer courses.

Stage Three

The Bard of Scotland, Robert Burns, gave us an oft-quoted phrase: “The best laid plans... gang aft agley.” The first of their plans to “go astray” appeared
when Principal Medwall said she could not arrive at a sensible logic model for the Gifted Program. The exercise had been valuable to her because it revealed just how much the program had grown without a clear vision. So far as she could tell, there was no agreed structure, and in her search she had consulted one or two experienced teachers over the summer. The teachers were also surprised that they could not arrive at a clear description of the program. All those directly involved had agreed to meet early in the fall term to begin afresh on a solid conception of the program (a ‘Proactive’ evaluation, according to Owen and Rogers). The Grade 10 transition program had a model of sorts, provided by the province, but Sheila had not been able to produce a clear description (with boxes and arrows), even with the help of summer instructors and fellow students.

The original professional development instructor was called in for an emergency consultation, and it was agreed that both teams should attempt to identify goals, objectives and outcomes, even before they were able to describe a plausible model. They probably knew where they wished to go, even if they were unclear about how they planned to get there – an outcomes-based evaluation had re-surfaced, under duress. Some evaluation criteria could be identified, and these were written down. Alas, there was no consensus how to measure/detect progress toward these criteria and the measurement specialist could not help in the absence of measurable objectives. The teams found themselves facing barriers that they could not pass, given the limited time they could devote to the tasks. Already they had given significant extra time to evaluation, believing that they could implement a useful version of SBE within the constraints of extra teaching load, new provincial requirements and reduced budgets for everything, including materials. They had to admit defeat, without reaching Stage Four.

After two years, a strong majority of teachers voted: (a) to cancel the formal SBE process; (b) to produce and circulate a written report; and (c) to continue with an annual report to the School Council and the Local authority on developments in the school and results on the EQAO tests. The general feeling, however, was that their own version of SBE had been an extremely positive experience, from which all the active staff members had learned a great deal. Everyone involved had a much clearer understanding of the complexities of their tasks and could see some positive steps that could be worked into the schedule in the coming years. The Mathematics Head retired on schedule, everyone enjoying a gala celebration in her honour.

**THE ROAD AHEAD?**

There seems little reason to believe that social and economic conditions will change in important ways in the near future. Neo-liberals have succeeded in
portraying globalisation and the unfettered free market as natural and inevitable, or at least vastly superior to other options. From a democratic, humanitarian and ecological point of view, however, neo-liberalism is a social pathology that if not checked will kill the organism in which it lives. John McMurtry (1999) outlines in detail an analogy between the human body, with its pathologies, immune systems and recovery mechanisms and the social body, society (in which many of us persist in believing, the Baroness Thatcher notwithstanding). McMurtry (1999) defines the "social immune system" as:

...an exactly articulated and regulated system of self and not-self recognition, continuous and comprehensive processes of surveying the social life-host for sites and phenomena of disease, injury and malfunction, and evolved organic structures and strategies of response to recognised impairments of the social body's vital functions (p. 89).

Program evaluation is one of the key "evolved organic structures and strategies of response" for the detection and healing of social malfunctions. Education, particularly public education, is a major sub-society within the full set of human relations we call society, and educational evaluation ought to be the backbone of our educational immune system. Educational tests and educational program evaluations could be vital organic structures and strategies of response to recognised impairments of the educational body's functions, and they were well on the way to attaining this status when they came into conflict with the neo-liberal agenda. As described above, school-based evaluation in Canada (the lack of it, that is) illustrates the point. SBE's highly evolved organic structure, combining local and external testing within a co-operative social and educational organisation could both recognise conditions that impair education's vital functions and identify positive strategies of response. Such strategies require time and money, however, and depend on a stable administrative structure that neo-liberals seek avidly to dismantle. As described above (the MSIP), external funding is not sufficient to generate an effective system to counter threats to the social immune system. A robust organic structure, e.g. SBE, is needed.

School-based evaluations and independent research studies are a threat to control by the common curriculums and large-scale testing favoured by neo-liberals. For example, charter schools and voucher programs are effective strategies for dividing the educational society into small units that are powerless to resist the overall economic agenda. Cynicism and disbelief are rational responses when neo-liberal "reforms" are justified in the name of efficiency and autonomy. At first, many teachers responded with enthusiasm to the freedom and collegiality in the best charter schools, but after only a year or two the full weight of governance, planning and evaluation are seen for the heavy burdens they are, burdens shared more widely among teachers and administration where
schools remain part of a local authority. Evidence is mounting that charter schools increase ethnic segregation (Cobb & Glass, 1999) and that the autonomy and security so often claimed are illusory (Bomotti et al., 1999). Neo-liberal ideology provides its adherents with their own immune system, however, making them fully resistant to evidence against their agenda. The agenda is, however, a disease agent, a cancer, within the social commons, including the educational commons (McMurtry, 1999, p. 115). Without resistance, it will destroy the social commons.

Signs of resistance are few. That school-based evaluation can be implemented in some countries is one of those signs, and the continued existence of independent research in universities can provide resistance to the cancer. That resistance occasionally breaks down, however, as in the case of University of Toronto Professor Nancy Olivieri. Dr. Olivieri was suspended from her research position in Toronto’s prestigious Hospital for Sick Children when she published a report critical of the drug she was studying without permission of the sponsor of the research – a drug company. Dr. Olivieri had unwisely signed a research contract that restricted publication, in contravention of University (but not Hospital) policy. She was reinstated after vigorous protests, but the stain remains (University of Toronto Bulletin, 1999). More recently, a formal offer of an academic appointment at the Centre for Addiction and Mental Health (affiliated with the University of Toronto) was withdrawn when it came to notice that Professor David Healy’s world fame was due to his pioneering research into potentially serious side effects of anti-depressant drugs, e.g. Prozac®. The drug company holding the patent to Prozac is a very large contributor to research at the University of Toronto, but everyone, including the President, denies that the links to the drug company had anything to do with the withdrawal of the offer of employment (University of Toronto Bulletin, 2001). The optimists among us, of which I am one, still believe that the immune system of the social commons will triumph because of the strength of the democratic institutions that have weathered wars and despots more than they have been forced to yield. For now, we fight a rear-guard action.

NOTES

1. Canada is much too large and complex for anyone to know all parts of it in detail. There may be examples of school-based evaluation; if so, they are few and isolated.
2. Halfway through the pilot program of this initiative the commercials were dropped, but the public service announcements were retained (McLean, 2000).
3. The Canadian private sector initiative stalled in fall 2000 when the major financial backer went bankrupt. As of summer 2001 it had not yet been revived.
4. Canadian legislatures, provincial and federal, follow the parliamentary system: the party winning the most seats forms the government. If they win more seats than all the others combined, they have a “majority” and can pass legislation without any agreement from the other parties. Because Ontario has three active parties, the Conservatives were able to gain a majority in the legislature with less than 45% of the popular vote. They call themselves “Progressive Conservatives,” but in my opinion the present version is more fascist than progressive.

5. This is a first-hand report from the teacher.


7. A good discussion and analysis of the concept of school choice may be found at http://epaa.asu.edu/epaa/v2n6.html

8. A comprehensive description and unbiased analysis of the charter school movement in the USA. may be found at http://epaa.asu.edu/epaa/v3n13/.

9. Secondary schools in Ontario were required by recent regulations to provide each student in several courses with a prescribed textbook and students in mathematics classes with a “graphing calculator,” at a total cost per student of about $400, even with volume discounts. Local authorities were given a designated grant for this purpose of $206 per student. Schools are expected to make up the shortfall of $20–30,000 from within their budget.


REFERENCES


INTRODUCTION

The Israeli educational system is characterized by a high level of centralization, and the evaluation mechanisms it presently employs are largely external, focusing, for the most part, on student achievement. Examples of these mechanisms are the National Matriculation Examinations (the Bagruth), which are used to assess the performance of graduating high school students, and the regional, standardized achievement tests for elementary school students.

Although these examinations are still widely used as the main form of student assessment, we have recently witnessed new directions in the evaluation of school quality. The catalyst for these changes can be traced to worldwide implementation of reforms in education, begun in the 1980s. These reforms were aimed mainly at restructuring the schools by constructing new forms of school organization and management (Murphy, 1992). This trend is reflected predominantly in the shift from bureaucratic and external control over schools to school empowerment: strengthening the control and influence of the school;
fostering principal leadership; promoting teacher professionalization (changing role definitions of principals and teachers) and strengthening the involvement of parents and the community (Hallinger, 1993; Murphy, 1992). With the expansion of school authority, schools were also expected to take greater responsibility for their actions.

In Israel, the ‘Autonomous School’ movement, initiated in the 1980s, constitutes an attempt by the Ministry of Education to implement such reforms. The ‘Autonomous School’ was defined as a school which sets its own educational aims, takes charge of the educational process and evaluates its actions (Reshef, 1987). Implementation of this approach can also be found in another program, the ‘School-Based Management’ project. ‘School-Based Management,’ an expansion of the ‘school autonomy’ concept, refers to enabling the principal and school staff to make school management decisions autonomously and to take responsibility for their actions (Friedman et al., 1997).

The components of accountability, responsibility and self-evaluation have been strongly emphasized in these projects. Thus, schools that participated in the ‘Autonomous School’ project were urged to develop internal mechanisms of self-evaluation. Prof. David Nevo of Tel-Aviv University endorsed and spearheaded the promotion of this approach, which became known as ‘school-based evaluation’ (Nevo, 1993a). The ‘School-based Management’ project also emphasized the importance of developing internal school evaluation mechanisms which would serve schools in their decision making processes and which would provide the basis for school accountability.

These new approaches to school management, teacher professionalization and school-based evaluation affected not only schools directly participating in these special projects, but other schools as well. In addition to the Ministry of Education, other intervention agencies have also been promoting the school-based evaluation approach, while implementing different programs in certain local educational systems and schools.

This case study will present one such experience, which took place in the town of Beit-Shemesh, Israel.

SCHOOL-BASED EVALUATION IN BEIT-SHEMESH

Over the past few years, numerous intervention agencies have engaged in ongoing intervention programs aimed at improving the level of the schools in Beit-Shemesh as well as the general effectiveness of the city’s local educational system. In cooperation with the local school authorities, one area chosen to be promoted at the community level was that of internal self-evaluation.
Incorporating internal, self-evaluation mechanisms into schools and into the local educational authority was perceived as a means of bringing about a change in the managerial perspectives held by principals and of contributing to the professional development of teachers. The goal was to create an “evaluation culture” or “evaluation minded-schools” (Nevo, 1993b) which would lead to global organizational change within the schools as well as within the local educational system as a whole.

In order to promote this idea, a steering committee was established in 1997. The committee was composed of representatives of the local educational system (including principals) and representatives of the different intervention agencies active locally. Mr. Avi Amram, the representative of the Keren Karev Foundation, who had expressed a special interest in advancing the idea, was appointed chairman of the steering committee.

The main guidelines decided upon by the committee were:

- Participation in the project was to be on a voluntary basis;
- Schools and the local educational authority were encouraged to choose an area/object for evaluation that was particularly relevant to them;
- Schools were asked not to evaluate student achievement in the initial stages of the project. This decision was based on a desire to expose the schools and the local educational authority to a broad range of alternative evaluation objects (special programs, curriculum, school climate, etc.), many of which were less frequently evaluated by the formal system.

In 1998, the project was implemented by an external intervention agency, hired especially for this purpose. Five schools as well as the local school authority decided to partake in the project. Each was assigned an evaluation consultant, recruited and employed by the intervention agency, to provide on-going guidance and training to internal evaluation teams. In addition, three community-based workshops were initiated by the program’s steering committee during the school year. These workshops were organized by the external intervention agency and were aimed at broadening the theoretical knowledge of the program’s participants: school principals, members of school evaluation teams, representatives of the local educational authority, school supervisors, as well as representatives of the steering committee.

Although there were some local successes, the implementation of the project, as a whole, was highly criticised by the general steering committee. These criticisms reflected a basic incompatibility between the conceptions held
by the steering committee and those held by the program implementers with regard to the nature of the program and its goals. The intervention agents perceived the project as an in-service training program, aimed at training teachers to perform internal evaluation in schools, whereas the steering committee and program planners viewed it as a broad-based intervention program, aimed at establishing formal internal evaluation mechanisms as a means of improving the overall organizational climate in the local educational institutions. Thus, the main focus of the project in its first year of implementation was on the conduct of evaluations of chosen topics. Questions regarding the establishment of an evaluation team as an integral part of school organization, definition of the roles of its members, and specification of the nature of the relationship between intervention team members, the principal, and other school personnel were not considered at all. Moreover, there was no clear conception of how school-based evaluation should be conducted within specific school frameworks. For example, in some schools the evaluation consultant worked with the entire school staff and an evaluation team was never formed. In another school, the evaluation consultant took on the role of program/curriculum developer and did not initiate any evaluation-related activities. Only two of the five participating schools experienced the evaluation process in its entirety.

Another criticism made by the steering committee was that the intervention agency had focused largely on individual schools, with little or no attention paid to the broader view of the school as part of the larger educational system. As such no genuine attempts had been made to encourage interactions between evaluation teams from the different schools, nor between them and the evaluation team of the local educational authority. Furthermore, little if any teamwork had been displayed amongst the project consultants themselves.

In view of the above, the work of the intervention agency was terminated after the first year of the project’s implementation, and the steering committee decided to take control of program implementation and its monitoring. In light of this decision, a special team was established to develop the concept of internal self-evaluation as well as a model for its implementation at the levels of the school and the local educational system.

PROGRAM DESCRIPTION AND IMPLEMENTATION

In the fall of 1998, the program’s development team convened for a series of brainstorming sessions aimed at outlining a preliminary model for the independent implementation of the program. At this point, the team included an academic advisor, the chairman of the steering committee, and an additional member of this body, who had been appointed to monitor and supervise the
professional aspects of implementation. Later, in-service consultants, whose role would be to implement the program in the schools and in the local school authority, were recruited and assigned to the program development team. All three consultants were professional evaluators, with some previous experience in evaluator training.

A preliminary draft of the implementation model, which defined the goals and conceptions underlying the program as well as broad guidelines for its operation, was presented in a series of “study” sessions attended by all the members of the development team. This draft was revised as the result of productive dialogue and brainstorming between team members. The program was ultimately implemented during the 1998–1999 school year by the in-service consultants, with the guidance and support of the program development team.

At the school level, the program adopted a participatory evaluation approach (Cousins & Whitmore, 1998). This approach constitutes the foundation for: (a) practical participatory evaluation, which supports decision-making and promotes organizational learning and change; and (b) transformative participatory evaluation, aimed at empowering people through participation in the process. Applying this approach at the school level was perceived as part of a broader framework of community-based goals which included the establishment of an evaluation “culture” at all levels of the local educational system, and promotion of dialogue at the community level, based on common knowledge and shared experiences in the area of evaluation.

As such, the program was perceived as a community-based intervention, comprised of multiple school-based intervention programs (see Fig. 1). Schools,
the local educational authority, or any other organizational framework involved in formal education in the community, were defined as target populations for participation in the program.

Five elementary schools opted to partake in the revised program, four of which had participated in the original program during the previous year, although only two had completed the prescribed evaluation process. The school authority entered the program but dropped out within the first few weeks of implementation.

PRELIMINARY MODEL OF IMPLEMENTATION

Our approach to developing an implementation model was one of flexibility. This is in keeping with the literature that indicates that each evaluation project seems to require its own tailor-made implementation model (see Alvik’s review, 1995, p. 320). General guidelines were set by the program development team and individual implementation programs were designed in cooperation with schools and in accordance with the needs as well as the constraints identified in each framework.

The final program consisted of two principal components: (I) A prescribed implementation framework for: (a) school-based evaluation, to be institutionalized over time so as to become an integral part of school life; (b) community-based evaluation, to be carried out at the level of the local educational system, and (c) an evaluation network generated through community-based workshops and study sessions for principals and evaluation teams from the participating schools, aimed at promoting dialogue between schools and sharing evaluation-related experiences and findings at the community level. (II) A broader implementation framework, including: (a) an in-service training and support system aimed at incorporating the above-mentioned framework into each institutional context as well as preparing selected staff members to take on the role of “internal evaluators;” (b) systematic monitoring and supervision of professional aspects of implementation; and (c) an internal framework for formative evaluation and reflection, designed to loosely monitor program implementation for the purpose of ongoing program development.

I. A PRESCRIBED IMPLEMENTATION FRAMEWORK

As previously outlined, our prescribed implementation framework relates to: (a) school-based evaluation; (b) community-based evaluation; and (c) promoting an evaluation network at the community level.
This framework consists of guidelines regarding implementation of the various aspects of evaluation. The following section presents these guidelines and describes their application in the field.

I(a) Implementation of School-based Evaluation

The Evaluation Team
Each school was expected to recruit a team of teachers who would be trained to engage in ongoing self-evaluation of selected aspects of school life (e.g. program, curriculum, and school processes).

1. Evaluation Team Composition: It was suggested that the team include at least two permanent members, one of whom would act as evaluation “coordinator.” Additional evaluation team members, whose expertise or interest may be relevant to the particular evaluation being carried out, could be recruited on an ad hoc basis. Principals were encouraged to choose charismatic teachers, ideally with some background in the field of evaluation, possibly holding key positions in the school and, most importantly, willing to commit themselves to this position over an extended period of time.

In practice, these guidelines were not strictly adhered to when recruiting evaluation teams. Albeit evaluation teams ranged in size from two to five members, however, coordinators were not chosen for any of them. This later proved to be a problem when tasks needed to be assigned, and responsibility taken for implementation of ongoing evaluation-related activities in the school. It should also be noted that two-member teams proved to be problematic, and post-implementation recommendations include increasing the requirement for evaluation team size to a minimum of three (preferably four or five) members.

2. Recommended Responsibilities of the Evaluation Team:
- Carrying out planned evaluations with the aid of additional school staff recruited for this purpose.
- Systematic self-monitoring of the evaluation process (evaluation journal).
- Organizing evaluation-related activities and workshops within the school.
- Involving relevant stakeholders (principals, general school staff, etc.) in the process at various stages of the evaluation (selection of topics/objects for evaluation, progress reports, etc.).
- Submitting a final written report to the program’s steering committee.
- Promoting dialogue and outlining recommendations in response to evaluation findings.
Evaluation teams in all the schools successfully completed at least one comprehensive evaluation program. Written reports describing the results were presented to school principals and to the chairman of the program steering committee. The remaining activities were carried out only partially in the different schools, depending on factors such as teachers’ workload and stress, attitudes and responses of principals and the wider staff to the process of evaluation taking place in their midst, and time allotted to evaluator-teachers for engagement in evaluation-related activities.

The Principal and His/Her Role in the Evaluation Process

A key issue in the successful integration of an ongoing evaluation framework in the school is the role played by the principal in this process. Within the confines of the present program, the principal was defined as the evaluation’s primary client; hence, it was recommended that he/she be actively involved in the evaluation process, though not necessarily as a member of the evaluation team. This involvement in and commitment to the evaluation process is seen as essential primarily because it indicates to all the school’s personnel the importance that the principal attaches to this issue. Moreover, this show of commitment and interest is necessary to maintain staff motivation throughout the evaluation process (Indrebo, in Alvik, 1995, p. 321).

This involvement includes:

- Participation in the selection of evaluation objects, and in defining evaluation questions.
- Support of the evaluation team throughout the evaluation process.
- Facilitation of the successful collection of data.
- Serious consideration of evaluation findings, provision of feedback to evaluation team members, and encouragement of dialogue relating to the findings, their interpretation and their implications.

Guidelines regarding the working relationship between members of the evaluation team and the principal were flexible, and recommended that consultations and on-going, verbal progress reports be incorporated into the evaluation process.

In practice, principals in the five schools displayed varying degrees of involvement in the process, depending, in part, on managerial styles and personality. In schools where principals had initially been extremely involved, there was sometimes a tendency to “step back” and allow evaluation teams to progress more independently. Here, it was sometimes necessary to re-involve principals at certain stages of the process so as not to ‘lose’ them and their
commitment to the program. In one school, the over-bearing nature of the principal’s involvement had a negative effect on teachers’ performance and had to be “neutralized” by the in-service consultant. Most principals showed differing degrees of involvement at different stages of the evaluation process, possibly the most “effective” form of involvement, assuming that they make themselves available to the evaluation team when their intervention is called for. None of the principals detached themselves totally from the process, and all, for the most part, were supportive of it.

Data from the teacher questionnaires administered to the members of school evaluation teams close to the end of the 1999 school year revealed the following: eight (of the twelve) teachers rated principal involvement in the process to be “adequate” while the remaining four expressed a desire for greater principal involvement.

Post implementation reflection has heightened our awareness of the centrality of the principal’s role in the evaluation process. Our current perception of the principal in the initial stages of implementation is one of a role model, an active member of the evaluation team, a leading force setting the tone for the entire school population in his/her endorsement of self-evaluation as an integral part of school life. The principal’s participation in the in-service evaluation training program is also essential for his/her own development as a “wise” consumer of evaluation. During the later stages of evaluation, the principal’s role is to support the evaluation process, to provide the conditions and resources necessary for its successful completion, and to seriously consider the evaluation findings and their implications.

To guarantee effective cooperation, the working relationship between members of the evaluation team and the principal should be clarified and institutionalized. Consultations, evaluation planning sessions and on-going progress reports should be pre-scheduled, when possible, in order to avoid scheduling difficulties later on.

Selecting Objects for Evaluation
Schools were given a broad set of options for choosing “acceptable” objects for evaluation and were free to determine both the nature and the scope of the evaluation: formative vs. summative evaluation; comprehensive as opposed to ad hoc evaluation of process and/or outcome variables; use of quantitative and/or qualitative evaluation methods; etc.

In their first year in the project, schools tended to choose objects which were seemingly neutral or less threatening: school climate, teacher/pupil satisfaction with various school-based enrichment programs, and the like. Of the “veteran” schools, only one evaluated a somewhat more sensitive issue – a program for
prevention of violence in the school. This same school had chosen to evaluate the "general school climate" during its first year in the program. In addition, schools tended to collect evaluation data mainly though questionnaires, and refrained from use of more direct methods, such as interviews and direct observations. Similar trends have been reported in the literature (see Alvik, 1995).

Post-implementation feedback from schools reveals a certain degree of openness to evaluating more sensitive areas of school life in the future (curriculum, teaching methods, principal/teacher performance and the like). Likewise, the expansion of data collection methods becomes relevant as schools gain experience and confidence in the area of evaluation.

The Role of the General School Staff in the Evaluation Process

Another factor left to the discretion of schools was the degree and nature of school staff participation in the evaluation process. Staff involvement varied from school to school, both in terms of its intensity and its scope. Schools in which staff involvement was more widespread were more inclined to engage in productive, evaluation-related dialogue, thus creating an environment conducive to the establishment of an "evaluation culture" in the long term. This was especially noticeable in one school, where the principal insisted that all staff members participate in a number of in-training workshops on evaluation. Unfortunately, in other cases, general staff involvement was minimal and even the task of recruiting occasional assistance from additional members of the staff (e.g. typing services, help with data collection) posed a major problem for evaluation team members and delayed the evaluation process. Post-implementation feedback (by means of questionnaires) revealed that six of the 12 members of the evaluation teams rated general staff involvement as "adequate" whereas the other 6 expressed a desire for greater staff involvement.

Post-implementation reflection increased our awareness of the great importance of general school staff involvement in the evaluation process. During the first year of the program's implementation, a series of introductory lectures was given to the general school staff to introduce them to the field of internal evaluation, and to impress upon them the benefits of self-evaluation in the school. However, the involvement of school personnel in the evaluation process should be more significant, especially in the early stages of the process, when the object of evaluation is selected and evaluation questions are formulated, later on, when assistance is required for data collection and finally, when findings are brought before the general school staff for consideration. Such an approach could promote the development of an 'evaluation culture' in the school.
I(b) Implementation of Community-Based Evaluation

As previously noted, the implementation of evaluation at the level of the local educational authority was terminated close to its beginning. Practical constraints and a certain degree of unease on the part of authority personnel at the prospect of introducing evaluation into the organization, prevented the program from "getting off the ground." An alternative to implementation of evaluation within the school authority was provided by school supervisors, who suggested that the 'English Language Program,' implemented by the Ministry of Education be evaluated at the community level. Although the prescribed model calls for self-evaluation of programs by internal evaluation teams, it was clear that these conditions would not be met due to the lack of prospective candidates for the evaluation team and the lack of time needed to organize a self-evaluation framework. Because we sensed the importance of carrying out some form of evaluation at the community level, an evaluation consultant was assigned to perform the evaluation in cooperation with the program's representatives.

I(c) Activities Aimed at Promoting Evaluation at the Community Level

(1) Principals and evaluation teams were invited to participate in community workshops and conferences as part of an agenda that sought to create group cooperation and evaluation networks so as to foster a sense of unity based on shared experiences and a common language, evaluation. These events also created an opportunity for dialogue amongst the schools, between the schools and the local educational authority, as well as with Ministry of Education supervisors, all of whom had a shared interest in promoting evaluation.

(2) Another means of promoting dialogue and cooperation at the community level was the inclusion of principals, a local school supervisor, and the head of the educational authority in the program steering committee. Their participation in this body, which met several times during the year, created an additional opportunity for discussion and collaboration.

(3) Plans to establish an evaluation network between evaluation teams and thereby to create cooperative learning situations, were aborted at this stage of the implementation, largely due to principals' and evaluation team members' resistance to this idea.

(4) A community newsletter describing the project and its activities was circulated at the end of the school year.
To summarize, it is apparent that in this case, expectations of creating a community-based "evaluation culture" may have been premature, and that pursuit of this goal will require a more extended process than originally anticipated.

II. THE BROADER IMPLEMENTATION FRAMEWORK

Establishing a basis for institutionalization of internal self-evaluation mechanisms at the different levels of the local educational system required the simultaneous implementation of various frameworks and procedures designed to meet the needs of program development. The following sections describe these supportive frameworks.

II(a) In-Service Training and Support Framework

Participating schools were each assigned a consultant, a professional evaluator, whose role it was to train the evaluation team to perform the various tasks required to complete a comprehensive evaluation program within the school. The underlying approach was one of teacher empowerment, the idea being that teachers become independent and creative internal evaluators, able to define evaluation questions, design evaluation tools and correctly summarize and interpret findings. Needless to say, this learning process is a lengthy one, as any reasonable level of professionalism cannot be achieved hurriedly. In some cases, schools felt that they no longer had the need for intensive guidance after one year in the program. However, their subsequent difficulty in performing evaluation independently convinced them that the process was more complex than they had previously believed, and that its success depended on the long-term commitment of principals, teachers and other school personnel involved in the evaluation.

The role of the in-service consultant was originally defined as one of "facilitator." The line between facilitation and active intervention or "taking over" is a fine one, and the challenge of remaining within the limits of facilitation was repeatedly addressed in group meetings between program consultants.

On average, consultants met with evaluation team members for in-service training once every two-three weeks. Unfortunately, as is common in on-going interventions, the process was hindered by cancellations (by teachers, for the most part), and the failure of evaluation team members to complete tasks assigned to be carried out between sessions. Heavy workloads and the lack of reasonable remuneration impacted negatively on teachers' motivation to persevere over lengthy periods of time. A special effort was made at different
stages of the process to meet deadlines and complete tasks but, for the most part, successful completion of the process depended on intensive involvement by the consultants.

In theory, as an "agent of intervention," the consultant was required to promote evaluation as an integral part of the educational process taking place in the school. This task included defining a working model of implementation, tailor-made to suit the needs of the individual school. An agreement was outlined with each school, which: (a) specified the latter's commitment to ensuring the allocation of resources necessary for the successful completion of the evaluation; and (b) outlined a clear framework for implementation (definition of roles, scheduled meeting times, specifications of work procedures, etc.). Schools were asked to sign this agreement at the onset of the program's implementation. In addition, consultants were encouraged to establish an advisory relationship with principals as part of an on-going process to instill a "culture of evaluation" in the school. However, not all principals were open to such consultant advances, and in some cases, the latter did not feel that they were able to establish a productive relationship with the principal.

II(b) Monitoring and Supervision Framework

Evaluation consultants met regularly with the representative of the steering committee appointed to the program development team as supervisor for the professional aspects of implementation. Within the framework of these meetings, implementation in each school was monitored, problems were addressed, potential solutions were discussed by the group (which thus served as a "support" group), and evaluation training materials were prepared collaboratively.

This group also met periodically with the program's academic advisor to plan collective training workshops for principals and evaluation teams (see section I(c)).

Consultants' verbal and written progress reports were included in the on-going monitoring and supervision. Detailed written reports were submitted by each consultant at the end of the school year.

Verbal contact was maintained throughout, between each consultant and the chairman of the program's steering committee, who supervised the administrative implementation of the program and monitored the process in the schools. Post-implementation meetings and individual interviews with consultants later revealed the need for input from an organizational advisor who would help consultants deal with the broader issue of implementing change in the school and institutionalizing evaluation mechanisms as integral part of the organization.
It became clear that the program required supervision not only in the area of evaluation but also in the area of program implementation. A decision was then made to hire the intermittent services of an organizational advisor during the coming school year.

II(c) An Internal Framework for Formative Evaluation and Reflection

Program implementation was accompanied by formative evaluation, a process aimed at identifying difficulties which might surface during the process of implementation, adjusting implementation guidelines when needed and assessing the program’s impact. Formative evaluation was conducted by the program’s academic adviser. The techniques employed included interviews with principals, questionnaires administered to evaluation team members, verbal and written program staff reports, discussions held during program staff meetings, and reports presented by participating schools at general program study sessions, workshops and steering committee meetings.

DIFFICULTIES IN IMPLEMENTATION

Many difficulties were encountered in the attempt to implement internal self-evaluation and to develop an ‘evaluation culture’ within schools and within the local educational authority of Beit-Shemesh. These difficulties can be classified as follows: (a) intervention-related difficulties which generally characterize implementation processes and arise in the course of most intervention programs; and (b) difficulties more closely related to the subject of our intervention: institutionalization of evaluation mechanisms in schools and at the level of the local educational system.

(a) General Implementation-Related Difficulties

- The program was initiated as the result of a decision made by community leaders to incorporate internal evaluation into the formal educational system of Beit-Shemesh. Thus, its implementation represents a ‘top-down’ approach to intervention, like many other programs and projects in Israel that have been initiated by the Ministry of Education or other external intervention agencies. The challenge here was to develop internal motivation, commitment and a sense of program ownership on the part of schools and the local educational authority. Other than emphasizing the benefits to be gained from the evaluation process itself and from its findings, attempts were made to foster the schools’ commitment by preparing a joint agreement defining the duties of the evaluation team, of the principal, and of other personnel involved
in the evaluation, and specifying the administrative procedures, allocation of school resources to evaluation, etc.

- Teachers' participation in the evaluation process increased their workloads and placed heavy demands on their time and on other school resources. The members of the Beit-Shemesh evaluation teams received minimal rewards for their participation in the program, so that much of their work was done on a voluntary basis. Clearly, these conditions cannot persist over time, and ways have to be found to reduce workloads and to properly reward the efforts of evaluation team members.

- Schools lack the organizational mechanisms to support the evaluation process; thus, they have difficulty providing the services and resources necessary to perform evaluation-related activities (secretarial services, computer services, teachers' working hours, etc.).

(b) Evaluation-Related Difficulties

Difficulties which were more specifically related to the implementation of the internal self-evaluation process corresponded with Alvik's (1995) account of some of the difficulties reported in the evaluation literature:

- Members of evaluation teams lack sufficient competence (knowledge and training) in the areas of evaluation and research methodology. The training process is lengthy and difficult; it takes considerable time to develop an understanding of what school-based evaluation is about, and longer still before the evaluation team can conduct a simple evaluation independently. Even after a lengthy training period, the process of data analysis continues to be especially difficult as it requires the use of computer-related skills. In light of these constraints, it is clear that evaluation teams need the extensive support of external consultants over a period of time sufficient to acquire the necessary knowledge and skills.

- Evaluation takes more time than expected and demands a great deal of effort from members of the evaluation team. Schools had difficulty setting aside the time required for internal self-evaluation and were frustrated by what they considered to be a low rate of productivity (the evaluation of one object took almost a year).

- Evaluation is often perceived as a threat. It can create unease amongst those implementing the program or project evaluated, or distress a principal who fears the loss of power and control. It is therefore important to emphasize the formative features of evaluation and to encourage the participation and collaboration of the entire school personnel throughout the implementation process so as to assuage any fears which may arise.
Together with the general difficulties which have been reported in evaluation literature, additional difficulties were encountered which reflect the unique nature of this particular program:

- As previously stated, the long-term objective was the development of an 'evaluation culture' which would affect individual schools, as well as the local educational system as a whole. The immediate goal was to develop internal evaluation teams within the school and the local educational authority which would serve principals/administrators as well as other members of these organizations. Experience has proven that this goal is not easily achieved: it is vital to first generate collaboration between members of the evaluation team, the principal, teachers and other personnel within these organizations. A considerable part of the evaluation consultant's efforts have to be devoted to the development of principal awareness with regard to internal self-evaluation and its benefits. Development of this awareness is essential for promotion of evaluation within the school framework and ensuring its meaningful use as an impetus for school improvement. It is also important that school personnel, other than evaluation team members and the principal/top administrator, be involved in the evaluation process. The entire school staff should be kept informed of the evaluation objects selected each year, of the evaluation findings, and of the decisions made on the basis of these findings. School staff may also be of assistance at different stages of evaluation (e.g. when formulating/revising evaluation questions and collecting data). As suggested in our model, school personnel can contribute more actively by joining the evaluation team during evaluation of the objects relevant to their areas of specialization. However, it should be clear that meeting the above conditions requires enormous input on the part of program implementers, including expertise in the areas of school reform, costly program support systems and well-organized systems for coordinating and monitoring the processes evolving in the field.

- The implementation of internal self-evaluation in Beit-Shemesh was intended as a community-based intervention at both the level of schools and of the local educational system. It was also meant to generate evaluation networks which would enable meaningful interactions between the various functionaries and would promote the concept of dialogue between internal and external evaluation (Nevo, 1995). The generation of collaboration between representatives of the participating organizations proved to be a difficult task. Three of the four principals who were interviewed did not attach any importance to such collaboration. Five of the eleven evaluation team members (in four of the five participating schools) stated, in response to a questionnaire item, that they were not at all interested in any collaboration with evaluation team
members from other schools. Three (of the eleven) were interested in some form of collaboration, mainly to hear what others were doing, although only three evaluation team members perceived such collaboration as "very important" as a means of learning from each another and of developing a sense of 'local pride.'

SOME EFFECTS AND BENEFITS OF INTERNAL SELF-EVALUATION – THE CASE OF BEIT-SHEMESH, ISRAEL

The 'internal self-evaluation' program has been operating in Beit-Shemesh for two years, but only in its second year was the program implemented according to a more comprehensive conceptual model. Change takes time. It is unrealistic to expect that significant changes would occur in the space of two years. At this stage of the project, however, we can present a few indications of some of the preliminary effects, based on empirical evidence derived from several sources: interviews with the principals; teacher questionnaires completed by twelve members of evaluation teams in four of five participating schools; reflections of evaluation consultants who worked closely with evaluation teams; descriptions of the evaluation processes experienced by the evaluation teams; and informal remarks made by various local educational system administrators.

• Improvement of decision making processes at the level of the school and of the local educational system

All the principals participating in the program reported that evaluation findings contributed to more effective decision making and substantially contributed to improved school processes and programs. They claimed that instead of operating on an exclusively intuitive basis, evaluation findings enabled more systematic thinking. As one principal put it: "When caught up in our work, we do things that we are not always sure are right. Evaluation provides us with the means of reviewing ourselves."

Several school-based programs were revised on the basis of evaluation findings. In one case, the scope of implementation was redefined while in another, program implementation was enriched by the addition of in-service training for teachers.

Members of the school evaluation teams expressed similar sentiments to those of principals. In reply to an open-ended question regarding the effects of 'school based evaluation,' eight of the twelve evaluation team members claimed that evaluation contributes mainly to improvement of school processes and programs, enables follow-up on changes that have been
implemented in the school, provides feedback on school activities and promotes self-examination.

The effects of evaluation were also felt at the level of the local educational system: On the basis of the evaluation findings, the content of in-service training for English teachers was revised to better meet their needs.

• **Experiencing the process of evaluation can serve as a means of promoting reflection**
Because the benefits of evaluation are not limited to its findings, as was pointed out by Shulha and Cousins (1997), the process of evaluation can be meaningful in itself. Participation in this process provides opportunities for reflection. As one principal stated; “We have learned how to ask questions.”

• **Internal self-evaluation empowers the individuals and organizations engaged in the evaluation process**
Cousins and Whitmore (1998) offer a new concept, “transformative participatory evaluation,” which they define as a type of participatory evaluation meant to “empower people through participation in the process of constructing and representing their own knowledge” (p. 8). The empowerment of schools and school personnel is an important goal in the broader context of school reform, decentralization and school restructuring, as was previously pointed out. According to Cousins and Whitmore, participation in the evaluation process contributes to the empowerment of an organization as a whole, in which case the school acquires the ability to monitor itself in a more systematic way and gains greater confidence in its educational direction. Participation in the evaluation process also contributes to individual empowerment, as members of the evaluation teams acquire new skills which they can apply in various contexts. The principals and teachers in our program reported that they were indeed able to apply the newly acquired evaluation skills and the knowledge of research methods in their schools. For example, teachers used their knowledge of research methodology to teach pupils to perform investigative tasks as part of their school work. Another example of such transfer of knowledge was the case in which teachers administered questionnaires to their pupils in order to assess student satisfaction with regard to various classroom matters.

The means of furthering teacher empowerment are not limited to those mentioned above. They may also include teacher participation in decision-making processes outside the classroom (Darling-Hammond, 1992; Hayman et al., 1995; Shedd & Bacharach, 1991; Wise, 1981), and fostering collegiality and collaboration amongst the teachers (Bacharach et al., 1986;
Darling-Hammond, 1992; Maeroff, 1988). As one in-service consultant reports, one noticeable effect of teachers’ involvement in evaluation was the formation of a pleasant working atmosphere amongst teachers of different subjects and different grades, within the school. This climate enabled mutual learning and fruitful collaboration. In addition, teachers’ participation in evaluation broadened their perception of the school as an organizational system.

- **Evaluation for accountability**
  As was previously emphasized, accountability and responsibility are central to self-management or to the ‘autonomous school’ approach. Schools which are granted a significant amount of authority, are required to be responsible and accountable for their actions. Although accountability was one of the goals of the ‘school-based evaluation’ program in Beit-Shemesh, it was decided that in the initial stages of the project, schools would not be required to report evaluation findings to external authorities. The schools’ staffs were expected to experience the process of evaluation, to gain confidence in their ability to perform evaluations, and to discover ways in which the evaluation process and its findings can benefit the school. Nonetheless, one school decided to report evaluation findings to parents and to present them with changes planned in response to these findings.

- **Evaluation as a basis for dialogue with external authorities**
  Internal self-evaluation is a necessary pre-condition for engaging in useful dialogue with external entities (Nevo, 1994, 1995). Nevo claims that “a school that does not have an internal mechanism for its self-evaluation will have difficulties in developing positive attitudes towards evaluation, and will lack the self-confidence necessary for a constructive dialogue between the school and the external evaluation” (Nevo, 1994, p. 96). Indeed, we have witnessed the cautious beginnings of such dialogue between the schools that participated in the program and the external authorities. One principal initiated a discussion with representatives of the Ministry of Education and Culture regarding the findings of an internal, school-based evaluation of a Ministry-operated program. Another example of such “dialogue” was a discussion between parents and the school evaluation team concerning the effects of a school-run program. As part of the evaluation in that school, parents were asked to fill out a questionnaire prepared by the evaluation team. The parents’ committee, in turn, decided to administer an additional questionnaire to the general parent body, which they composed themselves. The comparison of findings derived from each of the two questionnaires led to a joint discussion between parents and school personnel about the program and its effects.
Dialogue has also been initiated at a community level. The district director of the Ministry of Education and Culture initiated a discussion regarding the implementation of the local ‘English Language’ project, based on the evaluation findings.

- **Development of an ‘evaluation language’**
  A more general effect of the program was the development of an ‘evaluation language’ amongst program participants. The associated benefits were two-fold: Not only was improved communication witnessed amongst program participants, but familiarity with the language and concepts of evaluation enabled the promotion of a deeper understanding of evaluation and its impacts.

- **Additional uses of evaluation**
  The principals participating in the program have gone on to use evaluation in ways beyond those originally envisioned. In one such case, evaluation was used to promote public relations. During a presentation before representatives of the external authorities (e.g. ministry officials), the principal cited the existence of internal school-evaluation in his school as evidence of school quality. Evaluation was seemingly perceived as a means of enhancing the school’s reputation in the eyes of external functionaries. In another case, evaluation findings were used for political purposes: A principal who opposed the implementation of an external program in his school used evaluation findings to justify its termination.

**CONCLUDING REMARKS**

In our concluding remarks we would like to summarize the state of the program to date and to present our future vision of internal self-evaluation as an integral part of every school and of the local educational system in Beit-Shemesh.

The implementation of internal self-evaluation in schools and at the level of the local educational system is a process characterized by many difficulties. Some difficulties are general, arising during the implementation of any intervention program, while others are unique and directly related to the nature and content of the specific program, in this case, the development of evaluation mechanisms in organizations.

Despite the many difficulties which had to be overcome in a relatively short period of time, we have witnessed some signs of the program’s impact on the schools as well as on the local educational authority. The implementation of internal, self-evaluation in schools improved decision-making processes; served
The program was perceived by the schools, school supervisors and the local educational authority as important. Thus, all the schools which participated in the program during the 1998–1999 school year opted to continue with the project, even though this entailed some financial participation on their part. All eleven evaluation team members responding to our questionnaire thought that it was ‘important’ (five) or even ‘very important’ (six) to continue with the program in their schools.

Although at this stage of implementation the program has not had any significant impact at the level of the local educational system, the head of the educational authority and the local school supervisor representing the Ministry of Education and Culture attribute great importance to its continuation. Additional schools in Beit-Shemesh have shown an interest in the program and have participated in general evaluation workshops given at the community level. Parents in Beit-Shemesh have also changed their attitudes towards the program. Their skepticism has given way to support of program continuation and even expansion. Thus, the program was approved for the 1999–2000 school year by a large majority of the town’s general steering committee members, including representatives of the local educational system and of the parents. A growing number of stakeholders and educational authority officials have shown an interest in evaluation, and the foundations have been laid for evaluation-related dialogue within the schools, amongst schools and between schools, parents and the Ministry of Education.

A number of conclusions may be derived from experiences accumulated during the implementation of the program and from our reflections upon this process:

(1) It is essential that the principal be highly committed and involved at different stages of the evaluation process. Commitment and involvement are essential if the principal is to be able to stress the importance of evaluation, motivate the staff, ensure the availability of resources, and learn by this experience to be a ‘discerning’ client of evaluation.

(2) In order to institutionalize internal evaluation mechanisms in schools and to create an ‘evaluation culture,’ a much deeper involvement of the general school staff is necessary.

(3) It is important to broaden the range of objects chosen for evaluation and to include topics related to teaching and learning, even though these themes are more sensitive. It is also important to use a wider variety of
data collection procedures, incorporating more direct methods for information gathering, in addition to questionnaires.

(4) It would be preferable not to limit evaluation to only one object per year. It seems that schools' needs extend to a wide variety of evaluation-related activities; which could be carried out, ad hoc, in addition to a comprehensive evaluation of one major object. It is also important, for the long-term, to establish a school-based data pool and to analyze data collected at the school over time. This type of analysis would provide a comprehensive picture of the different aspects of the school and of the changes that take place over the years.

(5) During the initial stages of program implementation, evaluation teams need the intensive support of professional consultants. The role of the consultant is not limited to the training of evaluation teams but, rather, includes taking on responsibility for the institutionalization of internal self-evaluation mechanisms in the school.

(6) Special efforts have to be devoted to the development of evaluation mechanisms and evaluation networks at the level of the local educational system. The initial defensiveness of educational personnel and their unease with respect to evaluation must be dealt with, and genuine, collective collaboration must be fostered at different levels of the system.

The development of internal self-evaluation mechanisms in the schools and at the level of the local educational system in Beit-Shemesh should be considered within the broader context of ongoing trends towards decentralization of the national educational system and school empowerment. The function of internal, self-evaluation in the future will be to provide ongoing feedback for school improvement, to foster professional development and empowerment of teachers, and to serve as a means of meeting the demands for school accountability.

Internal self-evaluation can create "a solid ground for dialogue between the school and the external audiences interested in improvement" (Nevo, 1993, p. 4). This dialogue, based on evaluation findings, can promote mutual understanding, collaboration and a sense of partnership between the different audiences, be they teachers and parents, the community, the local educational authority, school supervisors and other Ministry of Education officials. Another dialogue which can develop is that between internal and external evaluation. As Nevo describes it: "It is a process in which at the beginning nobody knows everything, but both parties know something, and through the dialogue they know more and more" (Nevo, 1995, p. 189). Nevo, however, adds that in order to enable such dialogue, several conditions must be met. Most important, in our view, is the existence of "mutual respect and trust" between the parties.
Each party must respect the other professionally in order for there to be any meaningful discussion about evaluation findings. The parties must believe that this dialogue is in their mutual interest and that evaluation findings will be appropriately used. Unfortunately, such relations do not exist, as yet, between the different audiences in Beit-Shemesh. We still have far to go before our vision of a community in which internal, self-evaluation has become an integral part of schools and of the local educational authority, in which collaboration, mutual respect, trust and partnership have become the norm, and in which a constructive dialogue exists at all levels of the local educational system – and beyond – is realized.

REFERENCES


SCOTLAND: SCHOOLS SPEAKING FOR THEMSELVES

John MacBeath

INTRODUCTION

The British have long been wedded to the idea of external inspection of schools. In the opening years of a new Millennium the issue of school inspection continues to occupy the centre ground of debate nationally. While policy makers and politicians argue that this is the vehicle for raising standards, headteachers and teachers have voiced growing doubts. Could standards be raised by external pressure? Could quality be inspected into schools?

The four countries of the United Kingdom would probably answer these questions quite differently. They each have different systems of inspection, underpinned not only by different histories but by cultures (and languages) which bring with them different perspectives and value systems. The consequence is that while all four systems share some key similarities, there are there also significant differences in ethos and style, policy and practice.

The title given to England’s inspectorate, the ‘Office of Standards in Education’ (OFSTED) is an explicit statement of its role as the guardian of England’s schools – the popular notion of the ‘watchdog’ body peculiarly appropriate. The three assumptions which underpin that role are, one, that such a body or function will always be a necessary part of an education system; two, that quality can be inspected in the school and classroom life; and three, that...
an inspectorate provides an ‘objective’ view. These assumptions are explicitly stated in the following:

Independent inspection is here to stay. It is now an established and continuing part of school life. Inspection reports tell parents, governors and staff how good the schools is, that strengths and weaknesses are and what it should do to improve (Foreword, School Evaluation Matters, London, OFSTED, 1998).

The other three systems, all of which still retain Her Majesty’s Inspectors (HMI) take a less hard doctrinaire line. In Scotland none of the three assumptions suggested above would be seen as tenable. While there are some close parallels in the Scottish system with OFSTED there are also some crucial differences, the fundamental one being that schools should be allowed, encouraged and supported to tell their own story. While in the last few years OFSTED in England has progressively embraced school self-evaluation it has not been as deeply rooted in school practice as in Scotland and without a decade of gestation that has characterised the Scottish approach.

In the early 1990s the Inspectorate put in place a system of school self-evaluation in the belief that:

- schools should examine quality from different perspectives, in other words subjectivity counts;
- schools should take charge of their own evaluation and be furnished with tools to do so;
- in time, inspection (at least in its historic and current role) would wither away or be transformed.

In Scotland the road to mature self-evaluation is a long one and no-one could claim that, after nearly a decade, it is an integral highly valued feature of all Scottish schools. The spectrum of practice from the enthusiastically innovative to the grudgingly ritualistic is itself evidence of the slow bottom nature of the growth within a system.

THE EVOLUTIONARY REVOLUTION

The self-evaluation story starts at the end of a decade which has been described as the most turbulent decade in the history of the British schools. Under the Thatcher regime, a Secretary of State for Education, Kenneth Baker, was to bring in a raft of reforms which would radicalise the system and put into place ground-shaking policies whose aftershocks are still reverberating around the
Scotland: Schools Speaking for Themselves

245

...system and extend north of the broader; where, however separate the system, there is still a powerful Westminster Government to be reckoned with.

A watershed date was the 1988 Education Reform Act (England and Wales) which introduced school self-management. Bringing decision-making closer to the school could be presented as friendly to self-evaluation and educationally sound but that was not the overt or even covert, purpose of the reform. Its covert purpose was revealed by the architect of that system in a Guardian interview in late 1999. In the words of his interviewer (Davies, 1999):

The most seeping educational reforms this century, it transpires, had just as much to do with guesswork, personal whim and bare knuckle politics.

The Secretary of State’s self-confessed targets were first the teacher unions, removing their negotiating rights, then secondly the local authorities. Local school management was intended to fragment the teacher unions by giving them thousands of different employers to deal with and no chance of collective bargaining, and it would rob the LEAs [local education authorities] of their most powerful function by taking their hands out of the till ... He went on to give parents the right to choose their child’s school, thus robbing the LEAs of their second most important function, the allocation of pupils.

The covert intent, now explicitly acknowledged by Sir Kenneth, was to kill off the comprehensive schools and reintroduce selection, to create an open market which would see poorer schools (possibly in both senses of that term) being forced to close.

The overt intent, as expressed by the Department of Education itself, was

To put governing bodies and head teachers under the greater pressure of public accountability for better standards and to increase their freedom to respond to that pressure (Department for Education, 1992).

Under a Thatcher Government and a Thatcherite Secretary of State for Scotland, Michael Forsyth, these reforms were of more than incidental interest. Scotland had to steer its own course but tacking with the wind rather than trying to set sail against it.

Michael Forsyth, was prescient enough to recognise the solidity of comprehensive schools in Scotland (97% of Scottish pupils attend state comprehensives), the strength of the main teachers’ union, the (Educational Institute of Scotland), and the robust role of Scottish education authorities whose power and influence would not be undermined as easily as in England. He was a clever enough politician to take the advice of Her Majesty’s Inspectorate to adopt a more softly-softly approach and to follow a quite different course towards school self-evaluation and school improvement. Devolved school
management was introduced cautiously and pragmatically with regard to the historic "vertical partnership" (Raab, 1993) of government, authorities and schools.

SOWING THE SEEDS OF SELF-EVALUATION

In 1988 in Scotland the context for school improvement and accountability was external inspection. Although this did not mean that schools themselves were unconcerned with improvement and accountability, schools tended to see much of the responsibility for quality assurance and accountability as lying outside their control. Schools at that time could be characterised as:

• out of touch with current research and development in the fields of evaluation, and organisational development;
• seeing inspection by Her Majesty’s Inspectorate and/or the local education as a primary source for evaluating and validating of school quality;
• without the benefit of systematic criteria or tools for evaluating the quality of school culture, or organisational development.

Responsibility for this could not be placed at the door of the schools. It was a view deeply rooted in a wider culture and developmental history. At both local authority and national level there was a process which tended to reinforce the dependency of schools through inspection (both local authority and national) in which schools were largely passive recipients and criteria being used by the inspectorate were not disclosed to teachers or headteachers.

From 1988 onwards a number of things began to change. It could be described as a "sea change" in the sense of a slow swell, carrying beneath it an immensely powerful inertia. It was impelled in large part by what was happening on the United Kingdom stage but also because of mega-trends, international movements of idea. Scotland has benefited from an international outlook, as one Her Majesty's Chief Inspectors puts it, looking over the fence Impelling the sea change were numerous currents:

• a broadening international perspective – fact-finding missions, cultural exchanges, inter-country networking;
• a growing and widening debate on school quality and performance;
• publication of the HM Inspector of Schools school quality and performance criteria, opening up the Inspectorate's secret gardens;
• stronger emphasis on and policy-related research and competitive tendering;
• increasing reference by policy-makers to school effectiveness research and to school improvement;
• a strengthening of the researcher-policy-maker relationship.

All of these were constituent parts of a move towards a more effective systemic approach to school improvement. It implied a clear realignment of relationships between schools, authorities and national government. It was not the realignment that Kenneth Baker was pursuing in England where there a cutting loose of schools into a competitive self-sufficient market with strong monitoring from the top; but a new form of partnership, a new blend of internal school self-evaluation and external support and challenge. It was recognised that the marriage of internal and external evaluation is a matter of fine balance, one which many state or country systems have found difficult to achieve. It was acknowledged that it would take time to develop a new culture of trust and collaboration.

From the late eighties onwards teams of teachers, university researchers and Scottish Office policy advisers worked together to fashion a new system. In 1992 a set of self-evaluation guidelines were launched, distributed to all schools in Scotland, primary, secondary and special schools. They took the form of three large and luminously colored ring binders. Each contained an indicator framework, a set of suggested criteria, tools for self-evaluation, guidelines on their use, and examples of professional development activities for teachers. The Times Educational Supplement (September 15th, 1993) described this as “The Threefold Path to Enlightenment”. The three paths were:

• a set of qualitative indicators based on criteria used by HMI in inspections of classrooms, departments and aspects of whole school policy and practice;
• Relative Ratings, which gave secondary schools formulae for calculating the differential effectiveness of subject departments;
• Ethos indicators, providing schools with techniques for gauging the more subjective and least tangible aspects of school life from the perspectives of pupils, teachers and parents.

Taken together these provided schools with a set of procedures, a suite of indicators, both comprehensive and balanced, covering all aspects of school life, paying attention to the individual needs of least and most able, taking account of the personal and social as well as the academic, supporting and challenging teachers. It gave to schools for the first time the tools to conduct a full audit at every level of its operation or, alternatively, a selective progressive focus on specific areas of school life.
Although this was in some ways a hangover from the historic top down Scottish tradition, it was a significant milestone on the road to a more “school-owned” process of self-evaluation. One of the lessons to be learned over the next few years was that an ambitious dissemination of this kind can scatter too many seeds into the wind and that for good practice to take root green shoots of growth need to be tended and nurtured.

Where self-evaluation did take root and flourish quickly was in education authorities which supported experimentation, networking schools, sharing good ideas, providing extra resources and support to trial and test new approaches. For example, in one authority (Central Region Education Authority) a group of volunteer primary and secondary schools were brought together with the support from an external critical friend, to work together to pilot, develop and customise the indicators and test approaches to self-evaluation. One of the schools, in a depressed mining village in central Scotland, took the ethos indicators as a basis for further development. Rather than simply adopting the questionnaires and protocols provided the school took time to take a hard look at themselves, to create a climate of open reflection and collaborative investigation. Describing the school’s initiative in the journal Managing Schools Today (Ross, 1995), the headteacher outlined some of the key features of their eventual success:

- Before embarking on this self-evaluation process careful consideration was given to whether it was the right approach at that particular time.
- Management honored the professionalism of staff to be honest with them with no hidden agendas. Staff had to have ownership of the approach and agree with the questions to be asked of pupils and parents.
- Since the degree of openness and trust within a staff would determine the depth and sensitivity of the questions to be asked, management had to be particularly sensitive to existing morale.
- Anticipating surprising and perhaps disturbing elements the questions had to be faced ‘Are management and staff ‘big enough’ to take criticism?’ Are we prepared to ask the difficult questions?
- While guidelines provided a useful catalyst for thinking, questions had to be tailored to the school’s own context and the needs of its children.
- As school development was the end purpose, success rested on the care and quality of planning and climate setting.
- Account had to be taken of other events within the school calendar and other priorities within the school development plan. The ongoing work of the school could not stop. Effective learning and teaching had to be the priority with self-evaluation assimilated into it.
• Everyone in the school had to be clear about what was happening, when, and who was responsible for what.
• Time had to be set aside during in-service/staff development for decisions to be made about time scale, priority, and who should take the lead in planning.
• School assembly times were used to set the scene with pupils, with teachers following this up in class time, convincing pupils that the school genuinely wanted to hear their views, and that we would honor confidentiality. Pupils also had to know that what they said could directly affect future developments but guarding against raising their expectations inappropriately.
• Discussions with the School Board and P.T.A. were followed up with newsletters to parents. Parents' meetings provided a further opportunity to 'sell' the value and purpose of the initiative and reassuring them of confidentiality.
• Valuing the views of every pupil and parent it was important to cater for parents and children who had difficulty in reading. Questions were read to them as a whole class or in groups, according to age and ability. Those who couldn't read circled smiley faces. Those who could not cope with either were given a structured response sheet administered by teachers on a one-to-one basis.
• For parents with difficulties in reading volunteer parents played the role of helpers, working in pairs filling out questionnaires.
• Questionnaire were returned in a sealed anonymous envelope and a posting box was located in the school hall. On receipt of the questionnaire a raffle ticket was given. Personal contact and/or friendly letters were sent to those tardy in returning questionnaires. (producing a 100% response.)
• Quality time was set aside for staff to give considered responses, not colored by the stress of a long tiring day.

All of these proved to be important and formative principles, pioneered and shaped by teachers working together. The capacity to be innovative when given the opportunity and support is exemplified in the following passage from the headteacher's account:

A further opportunity to extend the discussion with parents was offered by the decision to photograph 'a day in the life of Coalsnaughton Primary'. John MacBeath, who was still working with the school, acted the proverbial 'eye' for the course of one day, providing a collection of photographs which gave a comprehensive insight into the ethos of the school. Here was visual evidence of relationships, pupil-pupil and teacher-pupil, giving us another focus to assess the ethos of the school and to deepen discussion on the school's main purposes and values. What did the photos say about the ethos we were trying to create?
How could that ethos support and enhance pupils' ability to learn? A second parents' evening was held. With the questionnaire results as a backdrop to discussion parents were asked to work in small groups, listing all the things they thought made a 'good' school. They then had to select and agree on ten. Their next task was to look through the photographs and see if they could find evidence in Coalsnaughton School to match their criteria of a 'good' school. Their final task was to prioritise these values by a pyramiding exercise putting the most important values at the top. This helped to broaden and contextualise the discussion of bullying and hitting back, getting us to the heart of the issue – the relationship between parents and teachers, home and school. In a small village such as Coalsnaughton keeping in touch should be easy for us all (Ross, p. 27).

Approaching the exercise as a learning, and risk-taking, opportunity the schools in this local authority project learned a great deal about themselves, about the school and about the process of self-evaluation. There were important lessons in it too for researchers, for local authorities, for policy-maker and for the Scottish Office.

**USING FEEDBACK FOR DEVELOPMENT AND ACTION PLANNING**

One of the most important of all lessons learned was the use of feedback of data as the bridge from evaluation to action and development planning. Exploration of data could serve, as in the Central Region schools, to stimulate lively discussion among parents, pupils and staff. An engaging aspects of these discussions was the comparison of teachers' concerns and priorities with those of parents or pupils. It not only created a new awareness but underlined the need for stronger more sustained exchange of ideas. It encouraged schools to be more accepting of diversity and conflict and more resilient in dealing with its implications. The term 'a culture of talk' has been used to describe school in which there is ongoing dialogue about where the school is going and why.

This process was explored more systematically and in greater depth a few years later in large scale project funded by the Scottish Office. The Improving School Effectiveness Project (MacBeath & Mortimore, 2001) ran from 1995 to 1998 and involved eighty primary and secondary schools in Scotland. A large body of attainment and attitudinal data was collected in each of these schools and fed back to the school as a basis for analysis, dialogue, planning and improvement. Twenty-four of the schools benefited from the support and challenge of a critical friend, appointed to work with management, teachers, parents or pupils in mining the data, exploring where it might lead.

The use of a double focus questionnaire, developed from earlier work in Canada (Stoll, 1995) required people to respond not only in terms of
agreement or satisfaction but in terms of expectation, so identifying a ‘gap measure’ between what ‘was’ and what ‘ought to be’. For example, in one inner-city secondary school the teacher questionnaire produced the following result:

<table>
<thead>
<tr>
<th>strongly agree</th>
<th>agree</th>
<th>disagree</th>
<th>strongly uncertain disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

Teachers in this school believe that all children can be successful

0 8 42 31 19

In terms of priority, however teachers said the following:

<table>
<thead>
<tr>
<th>crucial</th>
<th>important</th>
<th>quite important</th>
<th>not very important</th>
<th>not at all important</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

Teachers in this school believe that all children can be successful

78 22 0 0 0

What can a school do with such data? How important is it to have the help of a critical friend to effectively challenge a deeply rooted belief system? Could staff be convinced that closing the gap lay to a very large extent in their own hands? Could they come to see it not as a ‘problem’ but as an area of creative tension? At its very least the questionnaire data allowed schools to hold up a reflective mirror to their beliefs and expectations. It gave them the opportunity (“a first for us” as one longserving member of staff said) to engage in a collegial exploration of assumptions, of values, of words and meanings attached to words, moving towards the possibility of genuine dialogue.

As the relationship with the school continues and develops, the main contribution of the critical friend lies in helping people move to a more reflective dialogic approach, with an openness to questioning and respect for evidence. The question “how do you know?” eventually ceases to be put by the critical friend and becomes a routine way of thinking (MacBeath, 1998, p. 108).

WHERE TOP DOWN MEETS BOTTOM UP

The development of self-evaluation over a decade in Scotland illustrates that how top down meets bottom up, when two parallel processes are allowed to evolve. At the top level there has been an ongoing review and refinement
indicators and of instruments. The impact of guidelines and support materials resource packs has been evaluated and dissemination strategies reviewed. Universities and research centres have been funded to conduct ambitious research projects, to do independent reviews and to advise policy-makers on strategies for closer collaboration with schools. Funding has been provided by the Scottish Office (now reincarnated as the Scottish Executive) to establish dedicated centres in universities to serve as clearing houses, databases and honest brokers for inter-school networking.

Spotlighting and dissemination of good practice has been pursued through publications such as ‘Making it Happen in 12 Schools’, (Scottish Office Education and Industry Department, 1997), exemplary accounts of schools which had taken the initiative, often in the least propitious of circumstances and made things happen for themselves.

The policy story for the last half of the nineties is the progressive attempt to achieve the right blend between central guidance and direction and school-led evaluation and development. It is recognised that the role of external evaluators and inspectors should ultimately become one in which their assessment focuses on the rigour and integrity of the quality assurance mechanisms that are in place at school and classroom level.

**THE SLIM LINE APPROACH**

In 1996 the SOEID refined their national set of indicators bringing them down to a generic set of 33, applicable to primary, secondary and special education and to all groups within those three sectors. In 1997 “How Good is Our School?” was published and distributed to all schools, asking them to address three key questions for their own schools and to find their own answers:

1. How good is our school?
2. How do we know?
3. Now that we know what are we going to do about it?

The 33 criteria, or indicators, allow classroom teachers, departments and schools as a whole to evaluate their own effectiveness and progress, either using the criterion as formulated in the guidelines, or by using it as a model to develop a more personal statement. A simple four point scale, as used by Her Majesty’s Inspectorate, is suggested. The four points are:

- very good
- good – strengths outweigh weaknesses
Scotland: Schools Speaking for Themselves

- fair – some significant weaknesses
- unsatisfactory

There are qualitative descriptions to interpret what ‘very good’, or ‘fair’ would actually mean in practice. In respect of ‘quality of teacher-pupil dialogue’, for instance, classroom teachers might be asked to think through how they would describe high quality performance in that critical aspect of learning and teaching. They might then compare their own criteria with those offered in the guidelines. It is naturally tempting for a school to adopt these off-the-shelf criteria but teachers who have built on or customised the criteria have found it a valuable staff and curriculum development exercise. Thinking through the answers to these questions for yourself is much more likely to engender a belief in them and commitment to them. The essential principle here is that the process of systematic review, however it is undertaken, is more important than adherence to a predetermined structure or to one set of qualitative indicators.

The fact that these same criteria and this same four point scale is used by the Inspectorate means that when it comes to external review by HMI there is a basis for dialogue, a common framework from which to engage in joint evaluation of quality and standards.

What the processes exposes is not only the quality of the school but how well developed is the process of self-evaluation in the school. One member of the Scottish Inspectorate explained what can happen in schools where there is a superficial approach to self-evaluation (MacBeath & Myers, 1999).

It is quite common for a headteacher to run down the list of P.I.s and to say “We have done that. We have done that. So we should get a 4.” But the very act of seeing evaluation in this mechanistic way is in itself evidence of an inadequate approach to the issue.

At its best this collaborative pursuit of evidence tests the quality of thinking about management, leadership and effectiveness, and this forum for negotiation is seen by the Scottish Inspectorate as a seminal aspect of the quality assurance process, not only in arriving at some summative assessment but as a model in its own right for school self-evaluation (p. 198).

Schools and headteachers who are given a 1 or 2 in Inspectorate reports are in a fairly small minority, but Scottish HMI are at pains to emphasise the high stakes nature of that judgment and the care and scrupulous collection of evidence that needs to be gathered before such judgments are made. They are also keen to emphasise that the structure of the inspectorate with its internal accountability, monitoring and striving for consistency, is designed to inspire confidence in the quality assurance process. The question ‘How good is our Inspectorate?’ is pursued systematically along with the question ‘How good is our school?’
However strenuous the efforts of the inspectorate to combine fairness, rigour and consistency, its value will ultimately rest on the degree to which it is able to complement the school’s own internal evaluation and to strengthen the capacity of the school and school leadership to be self-evaluating and self-improving.

THE INTERNATIONAL DIMENSION

There is an international dimension to this too, both for Ministerial policy-making and for schools. They have to set their own performance and quality not just in a national context, but with a wider and more challenging vision. How well are they performing in a relation to schools in other countries? How much should we accept as inevitable and culturally determined? To help people explore these issues the Scottish Office has promoted international exchanges and funded international seminars. The Scottish Office Inspectorate played a significant role in the development of OECD indicators, leading one of the four international networks and developing a suite of indicators on public attitudes to education.

Scottish schools have taken part in a number of international projects on self-evaluation. Three Scottish schools are currently involved in the four year project funded by Germany’s Bertlesmann Foundation to develop and exchange innovative practice in school self-evaluation. Two schools participated in the European Commission’s ‘Evaluating Quality in School Evaluation’ (MacBeath, Schratz, Jakobsen & Meuret, 2001 and Michael Schratz’s chapter elsewhere in this volume). One outcome of the Bertelesmann Project was the translation was the adoption of ‘the Scottish Approach’ by a number of German schools and the translation of How Good is Our School into German – ‘Wie gut ist unsere Schule?’

In the context of the Scotsman newspaper’s national conference, addressed by both English and Scottish Chief Inspectors, Douglas Osler, Scotland’s Senior Chief remarked that in 1999 alone 31 countries had visited Scotland to examine self-evaluation both at policy level and as it works out on the ground in schools and classrooms. (Time Educational Supplement, September 10th, 1999, p. 3.)

In 1995 the National Union of Teachers (N.U.T.) in England and Wales commissioned a project to develop a self-evaluation framework in England and Wales following ‘the Scottish model’. Reprising the process that had been used in Scotland to develop ethos indicators, researchers worked with pupils, parents and teachers asking them to generate their indicators of what, for them, made a good school.
Scotland: Schools Speaking for Themselves

The research team was to discover that for many schools in the early 1990s these issues had never been explicitly addressed. Pupils, parents and teachers alike were taken aback, and delighted, to be asked for their views and surprised the research team with the acuity and perceptiveness of their insights. The process of asking them for a limited number of key indicators' helped implicit theories to be made explicit, assumptions to be tested and expectations open to mature reflection and debate.

Each group started from their own concerns. Parents wanted schools in which their children were happy and safe, learning good behaviour, how to get on with other pupils, to be treated with fairness and equality, to develop respect for others and be given the fullest opportunities to achieve and make the best choices for their future.

Pupils wanted schools in which they were treated fairly, with understanding and with firmness. They wished to be recognised as individuals and be listened to, their voice heard and their opinion considered without ridicule. They wanted teachers who could control the class, not allow those keen on learning to be penalised by the minority, but they wanted their teachers to have a sense of humour and make learning fun. They wanted to learn together, to work co-operatively. If they were in difficulty they wanted the teacher to recognise that and offer help.

Like parents and pupils, teachers started from their own concerns, with questions such as:

Is the school environment pleasant to work in?
Is there a climate of discipline?
Do you get the resources you need to do the job?
Do you get support from management?
Do you get support from parents?
Are decisions made with or without consultation?
Is staff development time used effectively?

These illustrate the different perspectives of those who teach and those who are taught, those who view the school from the outside through the experience of their children. and those who live day by day with its products, its failures and successes. It suggests that schools are satisfying or “effective” in different ways for different people and that the complex task for the organisation is to try and meet that range of expectations and to resolve the conflicts among them.

The more open and far-reaching the process of self-evaluation the more there is to discover about the value of these differing standpoints and different lenses through which the school could be viewed. When these stakeholder criteria
were put to the test through mechanisms such as interview, focus group, classroom observation or shadowing, people began to see their schools in new and unfamiliar ways. When pupils were able to see the classroom from the teacher’s viewpoint or when teachers saw the school from the pupil eye-view it opened a new world of understanding. When pupils had the opportunity to observe and classrooms and evaluate teaching their eyes were opened in new ways to the teacher perspective. When teachers shadowed pupils over the course of a day it gave them a more sensitive awareness of the pupil eye view. As one teacher reported:

I have to say it was a very mixed experience shadowing a pupil over the course of a day. It was like a ride over a bumpy road, sometimes quite painful, some jarring shocks to the system by some quiescent smooth passages too where little was happening, and some magic moments but, regrettably all too few. It was a very sobering reminder of what pupils experience day to day in our classrooms. It made me a much humbler and wiser man (unpublished teacher interview from the N.U.T study).

The N.U.T. study closely paralleled the experience of Scottish schools. It was, they reported, envigorating and challenging. The resultant self-evaluation guidelines were published in November 1995 as “Schools Speak for Themselves” and sent to all 45,000 schools in England and Wales and, as a follow-up study found in 1998, it the document been widely used as the basis for the development of local authority policies on self-evaluation (MacBeath, 1999).

**THE CHALLENGE AHEAD**

At national policy level in Scotland school improvement has been seen as a shared responsibility of people who make schools work from the inside and those who support and direct schools from the outside. The question ‘How good is our school?’ has both an internal and external focus. It has to be explicitly addressed within a school context and in a national and international context. Its point of reference is the larger question ‘How good is Scottish education?’ – one that has to be answered critically and with regard to the evidence.

The challenge ahead for Scottish education is to build on outstanding practice in the best of its schools. There will be voices and forces trying to push it back towards a safer more conventional model in which control rests in the hands of government and inspectors, in which as the OFSED document quoted at the beginning of this chapter, inspectors ‘tell’ schools how good they are.

Speaking on behalf of others, as Michael Fielding argues, not only defines their situation for them but, in a sense, even legislates who and what they are:
In speaking on behalf of others, Fielding continues:

I am engaging in the act of representing others' needs, goals, situation, and in fact who they are (Fielding, 1999, p. 5).

The task of the inspectorate may be seen as a parallel to that of the classroom teacher. As the teacher's task is to make herself as redundant as possible, so it is incumbent on the inspectorate. As the good teacher moves her charges progressively from dependence to independence so a wise inspectorate tries to break rather than to reinforce dependency. A wise system recognises that the significance locus of control and its powerful effects on organisations and individuals. Paul Martin has this to say about control:

It is hardly surprising that our minds — and those of other species — should be so attuned to a sense of personal control, since control over the immediate environment is vital for most organisms' survival. Control signifies autonomy, mastery and empowerment (Martin, 1997, p. 145).

Martin goes on to cite biological evidence as to the significance of personal control. He finds that:

- control or lack of it is closely related to mental and physical well-being;
- people with less power and control in organisations are more prone to heart disease than people with greater power and control;
- a sense of being in control helps to deal with and lessen stress;
- social support and networking lessens stress and risks of illness.

The perceived lack of control at both individual teacher and school level explain the stress so typically reported at or following OFSTED inspections. Fitz-Gibbon and Stephenson-Forster (1999) received responses from 61 schools recently inspected. The overall average for all 61 schools surveyed was 15.1 staff days before inspection. Absences during inspection dropped to 2.3 staff days on average, but then shot up again after inspection to 28 staff days on average. In terms of Martin's analysis this is a direct consequence of feeling out of control, being spoken for rather than speaking for yourself.

This is not a matter of either/or, not a matter of fruitless debate about external or internal evaluation. It does not invalidate the role of the external viewpoint or the enlightened eye (Eisner, 1991) but it does not pretend to objectivity. It brings, as the Japanese would say, the 'guest's view:'
The Japanese had no word for objectivity. It had to be invented once Westerners began talking. The word now used is kyakkanteki, literally 'the guest's point of view' (while shukanteki is 'the host's point of view' or 'subjectivity'. There is surely a hint of reproach in the first of the two terms. The guest is not simply an outsider and a stranger, but is probably naive about the nonvisible relationships of the family being visited. The guest 'sees' a number of separated people and, unlike the host, cannot know the pattern of dynamics within the whole (Hampden-Turner & Trompenaars, p. 113).

There is a similar concept in Icelandic – 'the visitor's side view'. This holds that the visitor to your house sees things that you never see because you have been become so accustomed to the everyday. Our sight is suffused with seeing what we already know rather than knowing what we are see, as Heschel (1970) put it. There is a peculiar quality to this external perspective. It acknowledges its own shortcomings and can, with the trust and confidence of friends, offer another perspective on reality.

In a mature system inspection can have that kind of quality. With a rigorous system of self-evaluation in place the role of the external agency is to ask not 'How good is your school?' but How good is self-evaluation in your school? What criteria do you employ? Where did they come from? How are they tested? What tools do you use? How valid and reliable is your evidence? To what extent is it built in to the fabric of the school and the say-to-day life of classrooms? What issues have been raised? What areas of school life do you still have to address? What kinds of evidence do you have to show that attest to quality and standards?

Self-evaluating schools seek, and find patterns of culture, mosaics of attitudes and expectations, motivations and achievements, ebbs and flows over time. They have a sophisticated understanding of the "reality" of school and the kaleidoscope of meanings for its various stakeholders. It is, as Hampden and Trompenaars (1993) suggest:

A culture attuned to the multiplicity of particulars or differences, in which it seeks to find patterns, may process information more easily than a culture searching for universal and uniform attributes... (p. 114).

The self-evaluating school has these characteristics:

- recognises that evaluation and school improvement is everybody’s business;
- has systems in place which engage all stakeholders in reflection and dialogue about quality and standards;
- embeds self-evaluation at every level of the school’s operation – classroom, department and school management and whole school;
- has a commitment to continuous improvement;
- welcomes external challenge;
Scotland: Schools Speaking for Themselves

• engages with different perceptions and sources of evidence about the school’s effectiveness.

That is the vision for the nation’s schools in the Third Millennium.

REFERENCES

INTRODUCTION

Although state school authorities have existed in Germany for nearly 200 years now, there has never been a tradition of systematic evaluation concerning school assessment or school quality in German schools. School development was always regarded as a case for global strategies, outlined in the ministry of education with curricula and school regulations and supervised by local or regional school authorities. But this kind of supervision or inspection is limited in a system where schools and teachers have a widespread freedom, as far as they fulfil the general objectives of their curricula.

Discussion about school quality and teaching assessment came up in the 1980s when research projects and interdisciplinary work-groups focused on the single school as the nucleus for school development and school quality. There was a change of paradigm from macro strategies to micro school politics on the one hand. But on the other hand international comparative studies (e.g. OECD) have shown, that German schools are rather mediocre in pupils achievement. So some German countries like Bavaria and Baden-Württemberg took these results to introduce standardized final assessments for A-Level exams. This was the first step towards comparative assessment in the sense of a country wide evaluation of academic student achievement. During the past couple of
years several countries have changed their school legislations introducing compulsory tests for the assessment of academic skills at various age levels from primary school up to high school. But most of these legislations have not come into practise yet. So there is little evidence about the outcomes of standardized tests concerning the increase of school quality.

Within this context evaluation and/or self-evaluation of schools has been promoted as a method for assessing school quality and bringing forward school development.

Without a unique experience with methods and instruments for school evaluation in Germany various models and objectives of evaluation have become relevant in the past ten years:

- evaluation in the sense of an actual diagnosis of the school as part of a school development process;
- evaluation as an assessment of the outcomes of the schools profile or programme;
- evaluation in the sense of quality assessment using performance or quality indicators for school subjects and teacher/pupil performance.

Concepts for self-evaluation are often combined with external evaluation as a new task for local and regional school authorities. External evaluators in this case should fulfil the role of „critical friends“ viewing schools achievement from an outer perspective and supporting school development with consultation and INSET-training.

THE PROBLEM OF DEFINING SCHOOL QUALITY

But the debate on school quality focussed more and more on the problem how to define quality and where to set standards and measure them. To measure the quality of products or services it seems to be necessary to have a common definition. For some products this is rather simple if you agree on durability, craftiness, or speed as quality indicators for products. The definition of school quality as a base for evaluation is far more complex. Some German school researchers even deny that it is possible to find a widely accepted definition (Steffens & Bargel, 1993). Others shift the problem when talking about various school qualities which compete (Posch & Altrichter, 1997). They argue that the definition of school quality is a process of social negotiation, where all participants (teachers, parents, students and politicians) should take part. In this sense school quality is a question of individual, ideological and societal status, influence and interest.
Following up the results of the TIMS-Study for German schools (Baumert et al., 1998) the quality criteria emerged as the cognitive skills of students in maths and science being measured with the help of standardized multiple choice tests. The ranking of assessment results in an international scale was the sole perception of the study in the public. The debate culminated therefore in the point, that the achieved rank in this "international competition" is the defining factor for school quality. The other results of the study stressing more the developmental aspects of the educational system like curriculum and teaching evaluation or the growth of student motivation in the teaching process did not seem to matter in the public dispute. Whereas critics of the study complained about the construction of the tests, which didn't fit into the teaching process and the curriculum in German schools, the supporters took the results as a proof for the lack of continuous academic assessment in the German school system.

One reason for limiting the definition of quality on academic achievement – especially by school authorities – might lie in the fact that assessing academic skills has a long tradition in schools. Every single teacher evaluates the skills of his students through tests and other more or less standardized instruments since generations. He probably doesn't compare his results with the ones of parallel classes or other schools. But the "micro cosmos" classroom works in similar ways as international comparative studies. Assessing academic achievement is a legitimate means to evaluate the "output" of schools, the "final school test scores". But though assessment is a legitimate and widespread practice in schools, the question if tests and measurements are valid to compare results in a ranking scheme is still rather doubtful.

**INDICATORS FOR SCHOOL QUALITY**

It is unquestioned that academic achievements of students are primary criteria for school quality. It is one of the most important output indicators of effective schooling. Test results and exams decide on the students' career and his or her chances on the labour market. But school quality is more than test results. This is one of the major outcomes of school effectiveness and school quality research of the 1980s (Fend, 1998; Scheerens, 1992). Output factors are highly influenced by input, process and context factors. A differentiation of evaluation indicators on these levels (see Table 1) are most important for the evaluation of schools, if not only the contents but also the goals of evaluation should be taken into account (Altrichter & Buhren, 1997).

School evaluation can aim towards the input factors of schooling, which are material resources or staffing, room situation or teacher recruitment or it can focus on context factors like catchment area, socio-economic status of pupils
Table 1. Quality Indicators on Input-, Process-, Output- and Context Level.

<table>
<thead>
<tr>
<th>Input</th>
<th>Process</th>
<th>Output</th>
</tr>
</thead>
</table>
| The main focuses of evaluation are the factors that are put into the system. These are for example the qualification of the teaching staff, the school resources (rooms, material, books, etc.), the curriculum and school legislations, norms and values set up by school authorities and the schools themselves. | These factors concentrate on school life and school culture, on the interaction between teacher and student and vice versa. They aim on the classroom activities, on teaching instruction, on school management or cooperation and communication in the teaching staff. Examples for process indicators are:  
  • Involving students in learning activities and decision making upon goals and curricula  
  • Instruction methods and how learning situations are arranged and organized  
  • Level of engagement in the teaching staff or the challenge of discrimination of minorities | The results of schooling, student achievements in a specific time schedule or effective school management concerning costs and resources are major factors of interest. The public image of the school and the satisfaction of students, teachers and parents are output factors of school evaluation. |
| If input indicators build the main focus of evaluation school quality is primarily seen as a problem of resources and the "right opportunity". | If process indicators build the main focus of evaluation the learning and instruction process as a context for student activities has a major value. | If output factors build the main focus schooling is primarily seen as a place for the production of effectiveness and efficiency. |

Context

Context factors affect the schooling process indirectly while emerging in the school environment. Normally these factors cannot be influenced by school development, but they can reflect the school climate whereas approaching chances and dangers can be extrapolated. Examples for context factors which are relevant for school evaluation and school development are trends of migration, changes in the labour market, youth unemployment, the development of new youth cultures, etc.

School Evaluation in Germany

These factors should nevertheless be ignored because they have a deep impact on schooling. But if school evaluation stresses the process of school development, it should concentrate on process and output factors (Buhren, Killus & Müller, 1998).

These factors are highly alterable and inventible in the single school. They underlie the change influence of teachers involving parents, students and school partners. In a complex sense process and output factors can be defined as the major items of a school programme. The school and classroom climate, the teaching and learning methods, cooperation and communication in the classroom and in the school, the management and organisation of schools, the development of specific norms and values are some examples.

School quality research has identified some of these factors as the major indicators for school quality (Tillmann, 1989; Steffens & Bargel, 1993; Posch & Altrichter, 1997). How to evaluate these aspects of school quality and school development is still not easily to be realized.

**MODEL PROJECTS FOR SCHOOL EVALUATION**

The actual practice of self-evaluation and/or (external) evaluation of schools in Germany are difficult to describe. With 16 German countries and 16 more or less different school systems and school ministries the extent up to which evaluation in one or the other form has already reached a practical stage differs very much.

In most cases evaluation practice in schools has not yet gone beyond an experimental level, though some countries like Berlin, Bremen, Hamburg, Hesse, Northrhine-Westphalia or Sleswig-Holstein have already changed their school legislation. In these countries evaluation is going to become a compulsory task for schools and school authorities. Schools are obliged to develop their own school programme, which should build the basis for internal and external evaluation. To illustrate evaluation practice in Germany, I would like to concentrate on Northrhine-Westphalia where research and pilot projects with schools and school authorities have started to develop methods, instruments and concepts for evaluation.

In 1992 the “Institute for School and Further Education” (LSW) set up a model project for schools and school authorities combining school development with formative and summative evaluation. Within a five years period schools took part in a development process – based on an organisational development concept with external consultants – experimenting with different evaluation instruments and methods like questionnaires, interviews, (self-) observation, etc. Members of the school authorities were working with the schools as consultants
as well as external evaluators. In most cases evaluation was used as an initial
diagnosis or state of the art of various school aspects (e.g. teacher co-operation
and communication, school management, school based curricula, management
of resources, etc.) heading for improvement and innovation. So far the project
has led to the following results:

• schools should be responsible for the process and the method of evaluation;
• evaluation should be embedded in a development process;
• participation of school authorities in evaluation can be critical;
• schools need accessible instruments for evaluation;
• schools need external support and training for evaluation.

A second research project on self-evaluation in schools was set up between
1996 and 1999, conducted by the Institute for School Development Research
(IFS). The project stressed the aspect of self-assessment so that the nine
participating schools had to present an a more or less elaborated school
programme as the basis for evaluation.

Evaluation in the participating schools concentrated on the following topics:
open learning, social learning, team building, job orientation programme, new
curricula, etc. Within the process of self-evaluation schools developed quality
indicators and quality criteria to assess their achievements. The indicators were
based on the schools' own objectives and the official curricula they were working
with. Within two years a cycle of evaluation was completed and evaluation
results were turned into a change of practice or the implementation of new
methods of teaching and learning. A main goal of the project was to develop
instruments for various fields of instruction and school programmes using
indicators to assess school quality and teaching/learning performances.

A follow up study of all participating teachers in the project schools made
clear that self-evaluation was widely accepted as a means to improve actual
teaching and/or school quality (Buhren, Killus & Müller, 2000) though teachers
complained about the amount of time they had to invest into evaluation without
receiving a lesson reduction from school authorities. The acceptance of school
self-evaluation can also be stated by the fact, that every single project school
set up new evaluation teams for other fields of evaluation after the project had
finished in September 1999.

Whereas all schools from the primary up to the vocational level had to
complete their individual school programme by the end of year 2000, school
authorities are now organising “evaluation dialogues” between the schools and
their school inspectorate. First experiences with this new form of external
evaluation indicate that quality and content of the school programmes differ to
School Evaluation in Germany

a wide extent. It reveals that schools still don’t have much competence to reflect and document their educational and pedagogical tasks and efforts or to plan for the future. But school inspectors also reported that schools which already had experiences with methods of self-evaluation developed more autonomous strategies to produce a pretentious school programme. This leads to the question how self-evaluation could be introduced to schools as a new means to improve school quality.

EVALUATION IN PRACTICE – A CASE STUDY

On the basis of an evaluation experience with about 15 different schools from primary school to high school in three countries the Institute for School Development Research (IFS) has developed a scheme for school evaluation, which divides the process into various phases (Buhren & Rolff, 1998; Buhren, Killus & Müller, 2001).

We call it the quality cycle of evaluation and it is mainly a process for self-evaluation of schools in nine steps. The first step is to choose the area of focus for evaluation. This can be a part of the school programme, a subject, a certain project of the school or the cooperation between school and parents. Choosing the field of evaluation is one of the most important steps in the process of evaluation. Here the school decides on the participants of the process and the amount of time and manpower they will spend on it. Schools should stick to the rule, that sometimes less is more. That means while planning their first evaluation project schools should have in mind their time and staff resources for the process.

A second step would lead to the main goals in the field of evaluation. What are the aims that the school combines with the field of evaluation or the project? What do they understand by quality or how is quality defined concerning this project? Is quality laid down in the curricula or in school legislation or is quality implicit in the self-defined goals of the school community (staff, parents)? Does quality emerge from the expectations of the public (industry, university, etc.)? Evaluation has to be linked to aims, if the success or the failure of a project wants to be measured. Otherwise the school goes through a kind of general diagnosis of their work, which could lead to aims in a follow up step. We have made the experience that schools often don’t take aims very seriously when they set up new projects or alter parts of their curricula. New teaching methods seem to be taken over because they are modern, but often teachers do not explicitly think about the goals these new methods should fulfill. So the definition of the main goals for some schools is a new way of looking at their daily routine.
The next step is to set up criteria for the main goals. That means to think about how the goals are going to be accomplished in practice. For example what are the criteria for more pupil orientation or for a mathematic course of high quality. If you have found the criteria you are able to set up indicators. The indicators value the performance of the criteria. So some schools call them performance indicators. We define indicators as items, which can be operationalized for data collection in surveys, questionnaires or observation studies.

Indicators will be formed into evaluation instruments in the fifth step. The data collection as the centre of the evaluation process should run through a pre-test. Our experience is that the “objects” of evaluation, the one who are going to be asked, observed or doing a self-assessment often give important information concerning the understanding and the practical value of the indicators. When the data collection is completed and the results on quality and quantity concerning the field of evaluation are compiled a feed back session with all participants in the evaluation process should follow. Especially when pupils or students have been the “objects” of evaluation they should take part in the feedback session to discuss and analyse the results.

Evaluation is not only the process of assessing an area of focus (e.g. a curricular area, topic, stage or teaching strategy). The nucleus of evaluation has to do with change and development. Therefore evaluation should lead to consequences, concrete targets a school is setting up, as a result of the evaluation process. There would be nothing more harmful for an evaluation process, if the data collection and the data analysis are left as goals in itself. Teachers would complain about the effort they have put into the process and a follow up evaluation would be difficult to set up. Targets of improvement can identify consequences of the evaluation process. The targets will lead to new aims in the area of focus or the field of evaluation. This will change school and teaching practice. An evaluation of the new practice can follow up in a couple of years. Then the cycle of evaluation starts again (see Fig. 1).

To demonstrate the cycle of evaluation in practice I will explain the different steps of evaluation with a small case study. It is the self-evaluation story of a grammar school in Northrhine-Westphalia using the evaluation cycle for their student oriented course work in three different curricular areas.

Choose the Field of Evaluation

A staff team at Newmarket Grammar School near Cologne has come up with the idea to evaluate their student oriented course work, they have been introducing five years ago in year group 7 and 8 in maths, social studies and
German. Their aim is to reflect systematically this concept, which includes new teaching methods, new teaching material and team teaching during certain periods of the total course time. They hope to gain data based information concerning the positive results and the failures of the concept, which they can analyse and value in order to plan alterations in their actual course work for a better teaching practice.

At the same time the evaluation process was headed towards some critics that have come up in the staff, expressed by a group of teachers who were not directly involved in the concept. They argued that their perception of the "alternative" teaching practice in these two-year groups was a rather poor achievement quality of the student's learning skills. They did not see any success of the project and doubted if the school would do well when they kept on working in the same way. The project time clearly realised that an evaluation would have it's pros and cons, because it could not only lead to alterations, but also to a complete stop of the project, if the estimated goals would not be achieved.

**Define Main Goals**

Projects can't work without commonly defined goals. Sometimes these goals shift out of sight, but they are absolutely necessary for any form of evaluation. Without having goals projects gain the character of self-fulfilling tasks. They may end up in daily routines where no one dares to challenge the actual practice.
They build a part of the school’s programme and this seems to be their only legitimate reason to exist. But when goals and aims of a project are clear and comprehensible they can lead to evaluation criteria and performance indicators. Then they would be able to value the positive and negative outcomes of the project. These were some of the thoughts and considerations the project team came to during their team sessions. They were trying to remember the goals they have set up five years ago, when the project was initiated and decided in the school conference.

*Develop and strengthen self-initiative and self-responsibility of students.*

*Increase motivation and individual learning skills of students.*

*Strengthen social skills like cooperation, solidarity, altruism, etc. of students.*

*Improve teamwork and group learning of students.*

*Develop individual learning abilities and learning strategies.*

These were the five main goals the project team has set up while introducing the concept to the school conference, but at that time they didn’t even think about ways and methods to evaluate if and how they would reach these targets.

**Set Up Criteria**

It is nearly impossible to evaluate all the main goals of a project, because they are often very complex and abstract. Therefore it is necessary to formulate and set up criteria that can be more easily measured with performance or quality indicators. A reduction to certain criteria might be helpful according to the fact that they should be able to be controlled directly by teachers. The project team at Newmarket Grammar School choose four criteria concerning their main goals: individual work, cooperation and teamwork, motivation and variety of methods.

**Set Up Indicators**

An operationalisation of aims and criteria is necessary if you want to measure them. That means indicators should reflect the practice referring to the major goals having been set up by the school or a project team. You can measure indicators by observation, description or questioning. The more detailed your indicators are the more information you get about the specific targets. Therefore indicators can help us assess the extent to which specific aims are being met. Aims provide general and specific expectations, which help to contextualise indicators by suggesting features to look for. Based on these considerations the
School Evaluation in Germany

271

project team was trying to find indicators for every single criterion. They set up the following indicators for individual work:

- Students work individually with material concerning the topic.
- Students contribute own ideas concerning the topic.
- Students try to find own solutions without the help of the teacher.
- Students develop own material concerning the topic.
- Students seek the help of other students if they have problems.
- Students work in their own pace without disturbing others.
- Students finish the tasks they have chosen.
- Students finish individual task with a written result.
- Students take over individual tasks.
- Students keep a folder where they document their work products.
- Students handle material and equipment carefully.

While setting up the indicators the project team realised that some indicators are relevant for two or more criteria. They found it difficult to decide whether to drop them if they appeared twice or three times. So they interpreted these indicators as the most important ones for evaluation and they understood that they would have to value some indicators different from others.

Setting up indicators is not an easy job to do if you have to find them for yourselves. They have to be related to your own school practice and you are not sure about their validity. The value of indicators for school or course work performance can only be noted when being used for evaluation several times. By comparing the results of your evaluation with previous results you will find out which indicators are helpful for future evaluations.

Collect Data

For the process of data collection the indicators have to be transformed into convenient instruments. For school evaluation there are more or less three ways of conducting data collections: You can observe teachers and pupils in classroom interaction and course work. So the indicators will provide as a scheme for observation. You can develop questionnaires where the indicators are formed into statements or questions on a scale base and pupils or teachers have to tick the most appropriate item scale as in multiple-choice tests. You can ask open questions in a questionnaire. The decision upon instruments for data collection depends on the number of persons being involved in the data collection and the amount of data you will probably receive with your instruments.
The Newmarket project team decided to work with two different forms of instruments:

- A standardized questionnaire being based upon the indicators.
- Classroom observation as a kind of participant observation of colleagues and self-observation of teachers where the indicators work as a kind of portfolio for the teachers.

The standardized questionnaire was given to all the teachers being involved in the project in form 7 and 8 to evaluate students' achievement concerning the set up criteria.

<table>
<thead>
<tr>
<th>Indicator applies for . . . % of students</th>
<th>100</th>
<th>75</th>
<th>50</th>
<th>25</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students work individually with material concerning the topic.</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students develop their own material concerning the topic.</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students bring in their own ideas.</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students handle material and equipment carefully.</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additional to that some colleagues agreed to being observed in their course work by members of the project team. The indicators formed the base of the observation accompanied by short descriptions of the teaching process, learning arrangements and the classroom interaction.

The perspective of evaluation so far didn't include the students' perspective concerning the project. That means while introducing all the different instruments of evaluation, the project team did not get any information how the students felt and thought about the concept of student oriented course work! It might be difficult to ask teenagers for a self-estimation on items like “I handle material and equipment with care” or “I try to find solutions without the help of the teacher”, because students have to look upon themselves in a distance. But without the students perspective the evaluation of the project is rather one-sided. The project team designed a simple questionnaire with just four open questions to cope with this problem:

- What do you like most in lessons with individual course work?
- What do you want to change in lessons with individual course work?
- What are your main difficulties in lessons with individual course work?
- What do you find difficult in lessons with individual course work?
Before using the evaluation instruments in all learning groups the project team decided to run a pre-test in two or three courses. The pre-test gave some information about items, which were unclear to teachers, so that alterations were necessary. Some items were dropped and others added. On the same time the project team could estimate the time and work effort, they would have to spend on the complete data collection.

On the whole 14 teachers in eight classes formed the basis for the evaluation at Newmarket Grammar School. Six observations of course work were added and 220 student questionnaires were returned to the project team.

Analyse and Interpret the Results

Evaluation results have to be matched against goals and expectations. Therefore you need a scale of measurement where you set up standards for quality. A 100% as the highest standard might be ideal to be reached, but this is an illusion in practice. The project team had to think about limits, where the assessment of the single items could be regarded as a success. They chose a 60% limit, that means if one item applies to more than 60% of the students in a course it could be defined as a strength. If the average of all indicators for one criterion dropped below the 50% limit they decided to define it as a weakness, so that a closer look at possible reasons and circumstances of results in this course would be recommended.

Give Feedback

When all the evaluation results where summed up and put into a report some members of the project team wanted to fix them at the pin board in the staff room. An information for the colleagues who were not directly involved in the project they considered it to be sufficient. Others argued that one reason to put up the evaluation process was the critics from that part of the staff, who did not actually work with the concept. So they made up their mind to present all the results in a special staff conference. Here they could not only discuss and interpret the results with all colleagues. They could also talk about consequences, which had to be drawn.

Draw Consequences

After nearly nine months of intensive work the project team was very content with their first evaluation process at Newmarket Grammar School. For the first time the staff conference discussed the concept of student oriented course work.
to the point based on data results. The analysis of the data made evident that no results were in the lowest scale (less than 25%). Improvements of the concept would have to be done in the area of cooperation and teamwork. Material for cooperative learning in course work should be developed to allow more effective group work instead of individual tasks for students. The evaluation instruments for this key area or criteria were to be used again in about six months to measure effects. The project team was asked to develop an instrument for students’ self-evaluation of course works similar to the standardized questionnaire for teachers. Course work observation was regarded as an effective means for cooperation between teachers and a self-reflection of teaching methods.

But the biggest success of the evaluation project was a decision of the staff to open up the concept of student oriented course work and include year groups 5 and 6. A couple of teachers wanted to get in contact with the three primary schools in the catchment area in order to get a closer look at the way these schools were working with similar concepts and to close up with their teaching methods when pupils changed to Newmarket Grammar School.

Install New Practice

The project team transferred the actual decisions of the conference into a project plan, which got a special place in the staff room – visible for everyone. Responsibilities for the installation of new practice at Newmarket Grammar School were laid out in the project plan and a time schedule should inform about the next steps of introducing student oriented course work in year group 5 and 6. Formative evaluation should value the different curricular methods and measure their success. Project management will be a new task for the project team. They will inform their colleagues on each staff conference about the project and thereby support transparency and cooperation.

SELF-EVALUATION VS. EXTERNAL EVALUATION

So far experiences with external evaluation have not gone beyond an experimental level in Germany. Whereas model projects of self-evaluation of schools seem to form an accepted basis for teachers, school authorities and school inspectorate still have a problem defining their role in the evaluation process. Perspectives and goals of evaluation differ when the main interest of evaluation lie on control and accountability instead of development. Figure 2 gives an impression of the underlying motives of self-evaluation and external evaluation.
School Evaluation in Germany

Fig. 2. School Development and Evaluation.
On the one hand evaluation can be seen as a process steering the school's development. In certain time periods the school will evaluate key areas of their work – for example based on an outlined school programme – in order to get detailed information about the performance quality and to improve these key areas. By undergoing a process of self-evaluation the main interest lies in the participation of the different groups (teachers, students, parents, community partners, etc.). The role of external evaluation in such a process can be defined as consultation. External evaluators try to understand and comprehend the inner school development. They seek for possibilities to support the school developing new projects or working on their school programme. External evaluators will look upon the self-evaluation as "critical friends".

On the other hand evaluation can be defined as a kind of diagnosis of the school's performance, to look at the strengths and weaknesses. The aspect of accountability could be a major concern of self-evaluation. A diagnosis can lead to further planning and more responsibility for the school development process and the achievements of the school. From the perspective of external evaluation diagnosis can also be taken for control. The school's performance will be assessed and ranked on the base of national standards. The output factors of schools play a major role in this process. External evaluators can use their power for regulation and top down strategies of school development.

Against this background of different interests, goals and perspectives concerning (self)-evaluation it seems to be necessary to find a common definition of evaluation inside and outside the school. This is the key for acceptance and a provision for the success of an evaluation process with or without external participation. The schools taking part in the model-project of self-evaluation have set up some aspects of a common definition, which could be a basis for understanding:

(\textbf{Self)-evaluation is . .})

- a systematic process;
- based on previously defined goals;
- leading to an assessment of a concrete practice;
- aiming towards the improvement and development of this practice;
- with commonly defined assessment grades;
- on data collection;
- with individually developed instruments;
- and a feedback for and with all participants;
- which includes concrete targets for the future.
If such an agreement upon the goals, the targets and steps towards evaluation has been found, different roles and tasks of all participants are easier to be outlined. Schools should not refrain from the possibility to involve external evaluators in the process, because they are able to take an outer sight on the school's performance – preferably as critical friends. In this sense members of the school inspectorate in Germany could take over a new role. School inspectors have a wide perception of the various activities and projects of the schools in their district. They know about the different input and context factors their schools have to cope with. They could support and consult the schools in the development of indicators or the area of focus. They should not interfere in the data collection and the analysis of results, but they could take part in the discussion and they planning of further developments.

Schools should be free to decide where and when they involve school inspectors in their evaluation process. And inspectors should respect the decision of their schools. An obligation for schools to work together with school authorities is – at this stage of evaluation practice in Germany – counter-productive concerning the overall intention of school development.

This is also relevant for other external partners from the school community or from neighbourhood schools. Although role and task differences will not appear in this case compared to school authorities, external participants especially from neighbourhood schools could provoke an atmosphere of competitive thinking. How are we doing in comparison to them! Individual interests may mix up with the targets of the school when partners from the community take over the role of an external evaluator.

School development and evaluation needs an atmosphere of respect and reliance. New forms of communication and feedback have to be set up. The dialogue between all participants forms the basis for development. This calls for an ethics and culture of evaluation to be established.

**ETHICS AND CULTURE OF EVALUATION**

The fact that on the one hand evaluation practice is still in it’s beginning in Germany but on the other hand school authorities have published several handouts and manuals where they unfold their expectations and general aims has led to some irritations among teachers and headmasters. The public discussion and debate upon evaluation has gone far beyond the actual experience with evaluation. The diversity of aims, the unclearness about the ways and principals of internal and external evaluation or a combination of both stimulates the arguments for and against evaluation.
It seems necessary to name principals of an ethic and culture of evaluation all participants can agree upon. These could be guidelines for action and common understanding (Buhren & Rolff, 1998).

Ethics of evaluation means in the first place that no one is to be blamed publicly concerning the results of an evaluation. Personal data may only be published if the person agrees with it. This is especially relevant when external evaluators take part in the process. The same accounts for school data, where schools have to decide whether they may be published and discussed in the public. School authorities should be careful with evaluation data of single schools. A ranking of achievement results, which can be read in the local newspaper, may serve the public interest but not the school.

Multi-perspectivity is another question of ethics. Schools should take the chance to involve external evaluators (parents, teachers from neighbourhood schools, etc.). In school research we use the term triangulation for a view on an area of focus from various sides.

All participants should accept the criteria for evaluation. That means that aims and targets of evaluation have to be set up, especially when school authorities or school inspectorate take part in the evaluation.

And last but not least feedback of results and achievements is a major aspect for an ethic of evaluation. Feedback supports dialogue and communication and creates a base for trust.

Establishing transparency and acceptance of the process can develop a culture of evaluation. There has to be reliability and the methods of evaluation should be transparent. Reliance has to be achieved through methods. It should not depend completely on persons. Aspects of a culture of evaluation:

- to anticipate the value of evaluation, that is to make clear whom it serves and what the aims are;
- to set up an evaluation committee, no one should evaluate alone. Evaluation is not the task of the school’s head team. They should be concerned with the setting up of a committee, they should care for the data control and for professional evaluation methods;
- to start small, that means not to evaluate the whole school, but to concentrate on certain areas, projects or targets;
- to use existing school data, that means schools produce regularly a number of input, output and process data, which can be embedded in an evaluation process;
- to write an evaluation report, where the process and the main results and outcomes of an evaluation are documented. Experiences have shown that reports of 15 pages have a reasonable and sufficient length;
School Evaluation in Germany

• to adapt or modify existing evaluation instruments for data collection. Schools don’t have to invent new instruments, they often can adjust existing ones for their own targets and evaluation interests. Cooperation with other schools might be helpful;
• to train the participants, for example in feedback methods and data collection;
• to publish the evaluation experiences. This can lead to more acceptance in other schools and to transparency in the public;
• to draw consequences, because otherwise the value of evaluation is difficult to conceive.

If evaluation of schools will support school development and increase school quality, self-evaluation is the best way to start with. Schools should initiate evaluation processes and they should be the main actors. But self-evaluation alone cannot secure school quality. Therefore it should be accompanied by external evaluation. To impose evaluation on schools by school authorities cannot be recommendable at this stage. For the next years evaluation of schools in Germany will go through a rather experimental phase. The more schools and authorities make good experiences with evaluation of one or the other form, the more it will be accepted and become a common practice.

REFERENCES
